CLASIFICACIÓN DE IMÁGENES DE ÁREA AMPLIA UTILIZANDO REDES NEURONALES CONVOLUCIONALES

Aplicación en Agricultura de Precisión

Ing. Oscar Andrés Martínez Silva

Proyecto de grado presentado como requisito parcial
para aspirar al título de Magíster en Ingeniería Eléctrica

Director
Germán Andrés Holguín Londoño M.Sc

Co-director
Mauricio Holguin Londoño, Ph.D

Grupo de Investigación en Gestión de
Sistemas Eléctricos, Electrónicos y Automáticos.

UNIVERSIDAD TECNOLÓGICA DE PEREIRA
MAESTRÍA EN INGENIERÍA ELÉCTRICA
PEREIRA
2021

Este trabajo está dedicado a mi esposa Lina Marcela, gracias por todo tu apoyo y creer en mi. Estoy seguro de que este logro no hubiese podido alcanzarlo sin tí, te amo.

También quiero dedicar este trabajo a mis hijos, ustedes son mi motor para continuar adelante y esforzarme por ser un ejemplo para ustedes, ¡los amo!.

Agradecimientos

En primer lugar agradezco a Dios, en quien creo, me guía en cada paso que doy. Así mismo, extiendo este agradecimiento a mi familia, que me han apoyado de forma incondicional, gracias por toda la comprensión y ayuda que he recibido y que me permiten culminar este proyecto.

También doy las gracias al grupo de Investigación en Gestión de Sistemas Eléctricos, Electrónicos y Automáticos, en cabeza del profesor Mauricio Holguín Londoño y a el director de este proyecto y mi tutor, el profesor Germán A. Holguín Londoño; sin su amistad, guianza, consejos y ayudas no hubiese sido posible terminar exitosamente este proyecto. Gracias por todos sus aportes, me han ayudado a crecer como profesional y como persona.

Quiero agradecer al ingeniero Jorge Luís Martínez Valencia quien me apoyó en la captura y creación de la base de datos de imágenes multiespectrales y el enlace con algunos agricultores que permitieron realizar los estudios sobre sus cultivos.

Por último, pero no menos importante, a la Vicerrectoría de Investigaciones, Innovaciones y Extensión de la Universidad Tecnológica de Pereira y al programa de Maestría en Ingeniería Eléctrica.

Este trabajo de grado es un producto resultado del proyecto investigativo denominado "UNA METODOLOGÍA PARA LA ESTIMACIÓN DEL ESTADO DE IRRIGACIÓN DE CULTIVOS DE ÁREA AMPLIA BASADA EN NDVI Y APRENDIZAJE PROFUNDO " adscrito a la Vicerrectoría de Investigaciones, Innovación y Extensión con código de proyecto 6-19-6.

A todos, mis más sinceros agradecimientos.

CONTENIDO

pág.

LISTA DE FIGURAS

pág.

LISTA DE TABLAS

pág.

1. INTRODUCCIÓN

Uno de los retos de cara al futuro de la humanidad es el abastecimiento alimenticio, el cual se ve afectado por factores como la explosión demográfica, la reducción de tierras disponibles para agricultura, el cambio climático, la escasez de recursos hídricos y el suministro de nutrientes para el crecimiento de los cultivos [1, 2]. En conjunto, estos factores hacen que sea necesario aplicar tecnologías con el fin de optimizar los procesos productivos en el interior de las granjas [3, 1].

Una disciplina que se encarga de estudiar las técnicas y la optimización de la producción agrícola es la Agricultura de Precisión. Con la aplicación de las técnicas y tecnologías de agricultura de precisión, se espera que los agricultores tomen decisiones acertadas en el cuidado de las plantas, permitiendo cosechas exitosas y optimizando el uso de los recursos a su disposición [4].

Dentro de las metodologías del estado del arte, sobresale el sensado remoto mediante imágenes para el análisis de grandes extensiones de tierra cultivada [2, 1]. Las imágenes pueden ser adquiridas mediante satélites, por medio de los cuales es posible observar grandes extensiones de tierra. Sin embargo, la resolución a nivel de suelo es baja y las imágenes se pueden ver ocluidas por nubes en zonas de la tierra con altos índices de precipitación, como es el caso del Eje Cafetero Colombiano. Estos problemas limitan la capacidad de análisis y pueden significar costos elevados cuando se trata del sensado periódico a determinadas zonas [1]. Otra forma de adquisición de imágenes, es mediante cámaras equipadas en sistemas aéreos no tripulados (UAS), con los cuales se puede aumentar la frecuencia de muestreo sin elevar considerablemente el costo de captura de datos. En comparación con los satélites, los UAS permiten obtener resoluciones en el orden de los centímetros por píxel, con lo que se facilita diferenciar un estudio de vegetación por especie cultivada en detalle [1]. La forma de analizar las imágenes

adquiridas por Drones consiste en el cálculo de índices de vegetación, los cuales, brindan información del estado de salud de las plantas, el estado de la irrigación e incluso la concentración de diferentes nutrientes [5]. Medir las variables mencionadas en un cultivo es posible gracias a que las plantas reflejan muy bien la luz verde e infrarroja cuando tienen una actividad foto-sintética alta, es decir, cuando son saludables y presentan un crecimiento apropiado [6, 7, 8, 9]. A pesar de que los índices de vegetación son una buena aproximación del estado de salud de las plantas, es necesario complementar esta información con sensores en tierra o con mediciones manuales, debido a que pueden ser ruidosas o inexactas. También la información adquirida como imágenes puede ser filtrada segmentando cultivos por especies y por zonas.

Adicionalmente hace falta de expertos en los cultivos que deben analizar una a una las imágenes, detallando cada una de las plantas para determinar los sectores donde se presentan afectaciones en la salud y por lo tanto muestran el crecimiento esperado [2]. Sin embargo, uno de los obstáculos al intentar estudiar conjuntos de imágenes aéreas, es que el volumen de datos hace difícil tanto la manipulación como la observación de fenómenos. Es por lo anterior que el cálculo de índices de vegetación debe ser complementado con técnicas de visión por computador y de aprendizaje de máquina, para refinar los resultados, facilitar el análisis y entregar mediciones confiables a partir de los datos adquiridos [10].

Autores como [10], muestran cómo es posible utilizar técnicas de aprendizaje de máquina profundo, para solucionar problemas de regresión o clasificación de especies en determinados cultivos, complementando la información adquirida por diferentes medios con máscaras que indican el tipo de plantas que se están observando. La clasificación de especies permite determinar y realizar un seguimiento al crecimiento de cultivos, su salud e incluso detectar de forma temprana anomalías como plagas o enfermedades que pudiesen afectan su desarrollo normal [3].

1.1. DEFINICIÓN DEL PROBLEMA

La agricultura de precisión es un término que agrupa el conjunto de metodologías y técnicas para optimizar los procesos productivos de las granjas [11, 12]. Cada especie a cultivar requiere mantener diversas variables en determinados rangos, como la temperatura, la cantidad de horas de exposición a la luz solar, la humedad relativa, el estado de irrigación y la concentración de nutrientes, con el fin obtener un crecimiento óptimo de cara a la extracción de cosechas.

Para poder medir cada una de estas variables, existen métricas [13] y conjuntos de sensores [2], que le permiten al agricultor tomar decisiones acertadas, las cuales, junto con visitas de expertos en la interpretación de los datos medidos, impactan en el estado de salud de las plantas, en la calidad del producto, y en la reducción del costo tanto de producción como de mantenimiento de los cultivos [14, 15, 16, 13]. Algunos autores [12, 17, 5] muestran cómo es posible conocer el estado de salud de un cultivo mediante fotografías con cámaras multi-espectrales, que dan la facultad de capturar imágenes con la reflectancia de la luz infrarroja en una escena, además de los canales de color que cualquier cámara del mercado es capaz de captar. La luz infrarroja es invisible al ojo humano, pero aporta información significativa en el trabajo con especies vegetales ya que estas, cuando tienen un buen estado de salud, presentan una actividad foto sintética constante, reflejan muy bien los componentes de luz en el espectro infrarrojo y verde [18]. Para visualizar y aprovechar la información lumínica reflejada por las plantas, existen índices de vegetación. Dentro de ellos se cuenta con el índice de vegetación normalizado (NDVI) [7], el índice de vegetación referenciado a tierra (GNDVI) [6], y el índice mejorado de vegetación (EVI) [6]. Cada uno de estos índices permite diferenciar en una escena, suelos, rocas, cuerpos de agua, plantas con baja irrigación y plantas con abundante irrigación.

Una de las dificultades para capturar imágenes multi-espectrales es el alto costo de los UAS con sensores infrarrojos integrados en el Drone y en el software de control de vuelo del fabricante. Como solución se puede anexar a un UAS una cámara especializada como el caso de la Sequoia Sentera™, que cuenta con lentes de captura en el espectro infrarrojo cercano. Sin embargo, para integrar la cámara con un Drone se requiere de personal especializado y soportes mecánicos a la medida, elevando el costo de tal implementación.

Una alternativa de bajo costo es el uso de un conjunto de dos cámaras como se muestra en [19], las cuales, permiten medir la reflectancia de la luz infrarroja sobre las plantas y la luz en el espectro visible. Esta solución implica tratar con dificultades que se pueden resolver por software como son: el ruido añadido por las vibraciones del UAS durante el vuelo, la distorsión radial de las cámaras, el desfase espacial entre lentes por la distancia focal de su disposición física y la sincronización de la captura de imágenes de los dos lentes [1].

Para eliminar el ruido por vibraciones y la distorsión radial es necesario procesar las imágenes. El proceso debe incluir la calibración de cámaras como se muestra en [20] y el ajuste de las imágenes por medio de filtros. El problema de la distancia focal entre los dos lentes también se puede solucionar calibrando las dos cámaras como un sistema estereoscópico.

Respecto a la sincronización de la captura de imágenes es posible integrar sensores en un mismo sistema embebido de control o utilizar georeferencias para determinar el lugar en el cual debe realizarse una captura. En caso de que la resolución de las cámaras sea diferente se puede detectar la zona de la imagen de mayor resolución que corresponde a la imagen de menor resolución [21, 19].

Otro factor importante en el análisis de tierras con imágenes aéreas, es la extensión cultivada que puede llegar a ser considerablemente grande, haciendo necesario tomar

cientos o incluso miles de fotografías para medir un solo cultivo. Tanta información implica el uso de dispositivos como servidores con almacenamiento distribuido, que a su vez requieren aplicar técnicas de de BigData para su manipulación y algoritmos de Deep Learning [10] para su análisis, que faciliten el procesamiento y entreguen información rápida y acertada a los agricultores.

1.2. JUSTIFICACIÓN

Según el censo agrícola de 2016, realizado por el Departamento Nacional de Estadística (DANE), la agricultura ocupa el segundo lugar en el PIB [22], siendo un componente importante de la economía del país. La extensión de tierras en Colombia es de 21 millones de hectáreas [22], se estima que 4 millones se utilizan para el cultivo de más de 95 tipos de frutas y cerca de 42 especies de hortalizas. En la última década las exportaciones colombianas de frutas han superado los USD 918 millones, equivalente a 1.83 millones de toneladas. La capacidad del país de producir diferentes especies radica en la diversidad climática, gracias a su ubicación geográfica y por la alta tasa de precipitaciones anuales (3.249mm/año) [22].

Teniendo en cuenta la ubicación de Colombia en un área tropical del planeta, las condiciones climáticas se ven significativamente afectadas por fenómenos naturales periódicos como las Oscilación del sur (ENSO), el fenómeno de El Niño y el fenómeno de La Niña, ocasionando según la época del año cambios bruscos tales como temporadas de sequía, donde administrar adecuadamente el recurso hídrico es fundamental para la conservación de los cultivos, así como temporadas de lluvias intensas donde pueden afectarse las cosechas debido al exceso de humedad [22].

Para administrar el recurso hídrico tradicionalmente se dispone de personal que debe recorrer el área, revisando el estado de las plantas. Como las extensiones de tierra son

amplias, esta técnica no siempre es efectiva, dado que es difícil mediante instrumentos tradicionales (palas y asadones) determinar el estado de irrigación de cada planta en particular. Una planta que tenga exceso de humedad es propensa a desnutrición por falta de nitrógeno en la tierra, el cual es desplazado por el agua. Así también, la falta de riego ocasiona muerte en los cultivos [2, 10, 23].

Dentro de las diversas metodologías para la medición y estimación del estado de irrigación, se tiene el uso de imágenes tomadas desde sistemas aéreos de bajo costo. Estos sistemas permiten proponer aplicaciones a la altura de las necesidades de los agricultores, que, a diferencia de técnicas como las imágenes satelitales, permiten tomar medidas en menores intervalos de tiempo y sin oclusiones por las nubes presentes. Las imágenes tomadas con sistemas aéreos no tripulados pueden ser utilizadas con índices que permiten saber el estado de irrigación, tal como el NDVI, el cual se puede estimar con imágenes que contengan información de la luz reflejada por las plantas en el espectro infrarrojo cercano [21].

En el caso del NDVI, se obtienen valores entre -1 y 1, asignando categorías a los datos encontrados según el número calculado para cada píxel [24]. La dependencia del índice respecto a la magnitud de refracción lumínica de cada especie de planta ocasiona que existan errores en la estimación, los cuales pueden generar clasificaciones inadecuadas del estado de irrigación de un cultivo. En el caso de un sector bien irrigado, si se clasifica de manera errónea y posterior a esto se vuelve a irrigar, genera exceso de humedad dañando las plantas [18]. Ese tipo de problemas se pueden solucionar aplicando técnicas de aprendizaje de máquina, en especial de aprendizaje profundo, debido a la cantidad de información que se obtiene al hacer tomar fotografías de grandes extensiones de tierra [18].

Es por todo lo descrito previamente que se evidencia la necesidad de desarrollar una metodología que permita la clasificación de imágenes de área amplia para la estimación

del estado de irrigación de cultivos, utilizando técnicas que que se aprovechen del uso de cámaras multi-espectrales, técnicas de visión por computador, procesamiento de imágenes, técnicas de aprendizaje profundo y agricultura de precisión.

El desarrollo de un sistema de segmentación junto con un dispositivo de bajo costo que permita estimar el estado de irrigación de cultivos de área amplia permitirá disminuir la brecha tecnológica existente entre los pequeños y medianos productores de la región de influencia de la Universidad Tecnológica de Pereira y aplicaciones que permiten mitigar el riesgo de la pérdida total o parcial de cultivos [5]. Además, permite allanar el terreno que deben recorrer los diferentes desarrolladores a nivel nacional e internacional en cuanto a la utilización de agricultura de precisión y aprendizaje profundo en la estimación del NDVI necesaria en diversas aplicaciones que hacen parte de la cotidianidad. Algunas aplicaciones que se pueden derivar del desarrollo de este dispositivo son: identificación de fugas de agua en sistemas de irrigación o abastecimiento, identificación de captación de agua no autorizada, análisis de la variación en la serie de tiempo del NDVI, seguimiento detallado de cada planta e incluso la estimación de la carga [19].

1.3. OBJETIVOS

1.3.1. Objetivo General

Desarrollar una metodología para la captura, gestión y clasificación de imágenes multi-espectrales de área amplia utilizando bases de datos relacionales y aprendizaje de máquina profundo (Deep Learning).

1.3.2. Objetivos específicos

1. Desarrollar una metodología para la adquisición de imágenes multi-espectrales utilizando un vehículo aéreo no tripulado.

2. Desarrollar un sistema de almacenamiento y gestión de imágenes multi-espectrales de área amplia.

3. Desarrollar un modelo de aprendizaje profundo para la clasificación de imágenes multi-espectrales de área amplia.

4. Desarrollar un modelo de aprendizaje profundo para la determinación del estado de irrigación de un cultivo

2. ESTADO DEL ARTE

2.1. COMPOSICIÓN DE IMÁGENES MULTI-ESPECTRALES

Una imagen multi-espectral es aquella que captura múltiples longitudes de onda del espectro electromagnético [25]. A diferencia de las imágenes convencionales, adiciona uno o más canales de color con información de radiación que no es visible para el ojo humano, como es el caso de la luz en el espectro electromagnético ultravioleta o infrarrojo.

De forma general, es necesario contar con múltiples sensores para adquirir una imagen multi-espectral [17]. Esta necesidad obedece a que la arquitectura de un lente convencional es similar a la que se muestra en la figura 1, en donde existe un filtro que impide el paso de la radiación infrarroja para que no se vean saturados los canales de color rojo y verde. La adquisición de una imagen multi-espectral se puede realizar disponiendo múltiples lentes con determinados filtros de color. La información obtenida por cada lente es almacenada de forma independiente y se evita la saturación. Sin embargo,la configuración descrita añade la necesidad de sincronizar espacialmente las capturas de todos los lentes, para que los objetos se aprecien en la misma posición en todas las capturas. De igual manera, es necesario compensar por software las diferencias inherentes a la fabricación de cada uno de los lentes como es el caso de la distorsión radial y las imperfecciones [19].

Una forma de realizar la compensación entre la distancia de los múltiples lentes, es utilizando un montaje rígido que impida el desplazamiento de una respecto a la otra. Posteriormente, mediante un patrón de calibración como se describe en [20], es posible obtener una matriz que define la calibración estereoscópica del sistema. La matriz descrita permite transformar las imágenes desde un plano proyectivo a otro, removiendo la

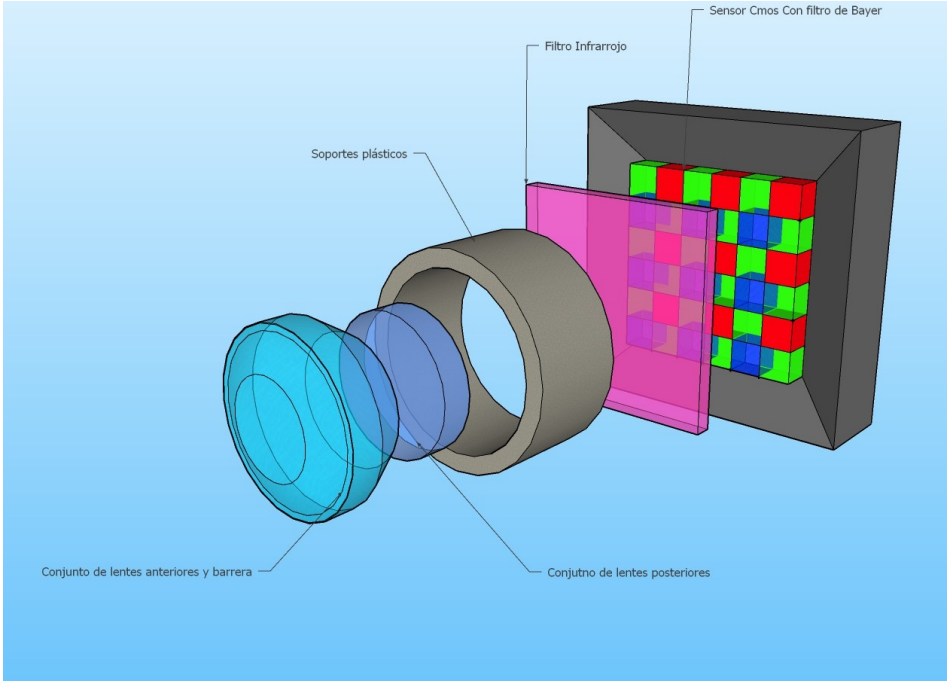


Figura 1. Arquitectura interna de un lente fotográfico estándar.

proyectividad, y por ende, corrigiendo el desfase espacial para que los objetos queden alineados en cada una de las diferentes capturas.

Una vez corregida la perspectiva se obtiene una alineación de imágenes con 4 o más canales de color. En el caso de las imágenes multi-espectrales para el seguimiento de cultivos interesa alinear principalmente los canales rojo, verde, azul e infrarrojo cercano. Con esta configuración se suelen calcular indicadores de vegetación como es el caso del NDVI , el cual muestra una aproximación del estado de irrigación de un cultivo [7].

Al tener al menos cuatro canales de color, una imagen multi-espectral no se puede almacenar en la mayoría de formatos conocidos. Tampoco es posible representarla en ninguna clase de monitor. Por ese motivo se almacenan los canales de forma independiente, con información que permita vincularlos y realizar cálculos para mostrar las composiciones en falso color, según la combinación lineal de los cuatro canales que se desee observar [21].

2.2. BASES DE DATOS DE GRAN ESCALA

La toma constante de imágenes multi-espectrales, está relacionada con la necesidad de almacenarlas, así como vincular cada una de las diferentes capas de color e identificarlas con etiquetas que permitan conocer su contenido o ubicación. Una vez almacenadas se requiere de cierto procesamiento para obtener información valiosa como es el estado de salud de las plantas. Debido a la naturaleza repetitiva de la actividad de toma de datos, toma poco tiempo llenar unidades de almacenamiento estándar con lo que hace falta utilizar múltiples unidades de almacenamiento externo a un sistema de cómputo [26].

Cuando el tamaño en memoria requerido por los datos supera las capacidades de los computadores convencionales e incluso de los servidores convencionales, se recurre al almacenamiento distribuido y al uso de técnicas de procesamiento conocidas como Big Data, que permiten organizar, vincular y analizar grandes cantidades de información guardada en servidores de almacenamiento distribuido o DataWarehouses [27].

También se necesita la implementación de algoritmos cuya complejidad sea acotada según la cantidad de datos disponibles, con el fin de evitar que una consulta tome grandes cantidades de tiempo. Existen motores de manejo de bases de datos con el fin de facilitar la creación, consulta y análisis de información. Dentro de estos motores se cuenta con los de tipo SQL o relacionales, que mantienen múltiples relaciones entre los datos facilitando la integración de información estructurada. También existen bases de datos no relacionales o NoSQL, cuya principal ventaja es la escalabilidad y soportan estructuras de almacenamiento distribuido. Cada una de estas alternativas depende del modelo de datos requerido por determinada aplicación [27].

2.2.1. Almacenamiento en motores tipo SQL

SQL es un motor de gestión de bases de datos centralizados ampliamente utilizado para todo tipo de aplicaciones que involucran la gestión de información. Dentro de sus ventajas se encuentra la facilidad de realizar consultas, mientras que el gestor de bases de datos preserva la información.

Con el paso del tiempo y el crecimiento de la información generada, las bases de datos de tipo SQL tuvieron que adecuarse para el problema de Big Data, donde no es posible almacenar todos los datos en un único servidor, sino que se necesitan múltiples servidores. Esto añade problemas de sincronización y de retardos cuando una consulta es ejecutada.

Para poder gestionar ese tipo de bases de datos se desarrollaron tecnologías basadas en SQL como SQL on Hadoop, con las cuales se pueden realizar consultas en servidores de almacenamiento distribuido.

2.2.2. Motores de big data (HIVE)

HIVE es un entorno para la gestión de bases de datos con grandes cantidades de información, desarrollada por la fundación Apache basados en el paradigma de bases de datos no estructuradas, del tipo NoSQL. Dentro de sus ventajas se encuentra el almacenamiento distribuido en diferentes servidores de datos, la confiabilidad en las consultas de datos, y la escalabilidad de sistemas paralelos. Este entorno permite manipular bases de datos con información estructurada y no estructurada en archivos muy grandes almacenados en múltiples servidores.

La plataforma consta de: (1) Disponibilidad de los datos mediante el sistema de almacenamiento distribuido de Hadoop (HDFS). (2) Estructura YARN de Hadoop para

planear tareas y manejar clusters. Y (3) Implementación de MapReduce basado en la estructura YARN para el procesamiento en paralelo de conjuntos de datos grandes [28].

Una de las ventajas radica en el manejo de grandes cantidades de datos, los cuales pueden manipularse en paralelo para reducir el tiempo de cómputo aprovechando los clusters de servidores, lo cual permite estimar índices como NDVI optimizando el tiempo necesario para tareas de análisis de extensiones de tierra amplias [28, 29].

2.3. SEGMENTACIÓN SEMÁNTICA CON REDES PROFUNDAS

La segmentación semántica consiste en tomar cada uno de los píxeles que componen una imagen, y asignarles de forma objetiva una etiqueta, con la cual se identifica a qué categoría pertenecen. Por ejemplo, en la figura 2 se puede apreciar cómo una imagen, luego de pasar por un proceso de segmentación semántica, obtiene etiquetas para los píxeles pertenecientes a la clase fondo (color negro), bicicleta (color verde) y persona (color Rosa).

Figura 2. Ejemplo de segmentación semántica en la base de datos PASCAL VOC [30].

La segmentación semántica es posible gracias a los resultados de las redes convolucionales profundas, cuyas arquitecturas han demostrado estar en la capacidad de resolver

tareas de detección a nivel de píxel. Una de las redes pioneras en segmentación semánticas fue AlexNet que, en 2012 reportó una precisión del 84.6 % con una arquitectura de 5 capas convolucionales [31]. En 2013 la red ganadora de la competencia ImageNet fue vgg-16, que reportó una precisión del 93.3 % con una composición de 22 capas y un bloque denominado *inception* [32]. Posteriormente uno de los aportes más significativos en detección a nivel de píxeles y en profundidad de la red fue el equipo de *Microsoft* quienes propusieron la arquitectura denominada Resnet [33]. Este modelo logró un 96.4 % de precisión con una red que alcanzaba 152 capas de profundidad. Una de las dificultades de las redes profundas era el desvanecimiento del gradiente cuando se apilan muchas capas convolucionales sucesivas. En Resnet el desvanecimiento fue controlado gracias a la inclusión de una arquitectura denominada bloque residual. La principal característica de los bloques residuales son sus conexiones "de salto" llamadas *Skip connections* y permiten conectar la última capa de un bloque con la suma de las entradas ponderadas y las salidas de las capas convolucionales. Esta técnica permite reducir el condicionamiento numérico del gradiente, abriendo paso a arquitecturas mucho más profundas que exhiben mejores resultados en tareas de aprendizaje de máquina.

2.4. ARQUITECTURA *U-NET*

U-net es una arquitectura de red profunda desarrollada en el centro para estudios de señales biológicas de la universidad de *Freibrg*, Alemania, en el año 2015. Fue utilizada originalmente para segmentar células en imágenes biomédicas de un canal de color. En la publicación [34] los autores superaron con un margen considerable al mejor algoritmo para resolver el reto ISBI en la categoría de segmentar estructuras de células neuronales en imágenes tomadas por microscopios electrónicos como se muestra en la tabla 1.

U-net ofrece múltiples ventajas tales como la velocidad de inferencia, que es inferior a un segundo; la implementación completa, que puede ser ejecutada con recursos de GPU moderados y el entrenamiento, que puede ser realizado con pocas imágenes.

Para medir el desempeño de *U-net* los autores utilizan la métrica de intersección sobre la unión, también conocida como IoU o Coeficiente de Jaccard. Esta métrica se calcula como se muestra en la ecuación (1) y representa la coincidencia entre las zonas detectadas por el algoritmo en comparación con las etiquetas realizadas manualmente por humanos. Los resultados de *U-net* respecto a IoU se aprecian en la tabla 1, donde también se relacionan los conjuntos de datos que utilizaron en [34] para el entrenamiento, el número de imágenes parcialmente etiquetadas y el rendimiento de la red. En la última columna se aprecia el rendimiento del algoritmo que mejor desempeño presentaba antes de *U-net*. Se evidencia que, a pesar de tener pocas imágenes el rendimiento es bueno y es significativamente mayor a técnicas como redes neuronales con ventanas deslizantes, que dominaban el reto ISBI antes de 2015.

Dataset	Número de imágenes	IoU Unet[%]	IoU Algoritmo anterior [%]
PhC-373	35	92.03	83
DIC-Hela	20	77.56	46

Tabla 1. Resultados de *U-net* reportados en [34].

$$IoU = \frac{AreaTraslape}{AreaUnion} \tag{1}$$

U-net es una arquitectura de red profunda compuesta por tres secciones, una de contracción, el cuello de botella y una sección de expansión. En la primera sección, que también se suele llamar *encoder* existen varios bloques que aplican filtros convolucionales de 3×3 y posteriormente tiene una capa de *max pooling* de tamaño 2×2. Con cada bloque se duplica la cantidad de canales en el tensor, equivalentes a mapas de

características. De esta forma se puede aprender relaciones complejas y estructuras de objetos en una imagen [34]. La segunda sección propaga las características con capas convolucionales, esto con el fin de realizar una extracción densa de características. En la tercera sección, también conocida como *decoder*, se presentan bloques de expansión similares a los de la primera sección, que contienen capas convolucionales con filtros de tamaño 3×3 seguidos de una capa de sobre muestreo o *upsampling* de tamaño 2×2 que aumenta el tamaño del tensor y disminuye el número de canales. De esta forma cada mapa de características mantiene una simetría en tamaño con respecto a los bloques de la primera sección. Esta simetría permite interconectar los bloques de la sección de contracción con la sección de expansión, garantizando que las características que se aprenden mientras se contrae una imagen, son utilizadas para la reconstrucción de una máscara de segmentación. De igual forma esta técnica de interconexiones evita problemas como el desvanecimiento del gradiente [34].

Suponiendo que la entrada de la red es una imagen de tres canales de color con una altura H y un ancho W de píxeles, es decir, un tensor de dimensiones $H \times W \times 3$, en la salida de la sección de expansión de *U-net* se obtendrá un tensor de dimensiones $H \times W \times k$. Siendo k el número de clases de objetos a segmentar en la imagen, en otras palabras, cada clase genera una capa donde se identifica por cada píxel la probabilidad de que pertenezca a la k-ésima clase. A partir de ese tensor, se puede lograr una máscara de segmentación, es decir, una imagen de un canal con dimensiones $H \times W \times 1$ aplicando una función SOFTMAX a lo largo de las k capas como se muestra en la ecuación 2. Esta función es una exponencial normalizada y permite comprimir las K capas en una sola capa $\sigma(Z)$. A cada píxel se le asigna un número entre 0 y $K-1$ según el valor que mayor probabilidad acumule sobre las K capas [35].

$$\sigma(Z)_j = \frac{e^{Zj}}{\sum_{k=1}^{K} e^{Zk}} \tag{2}$$

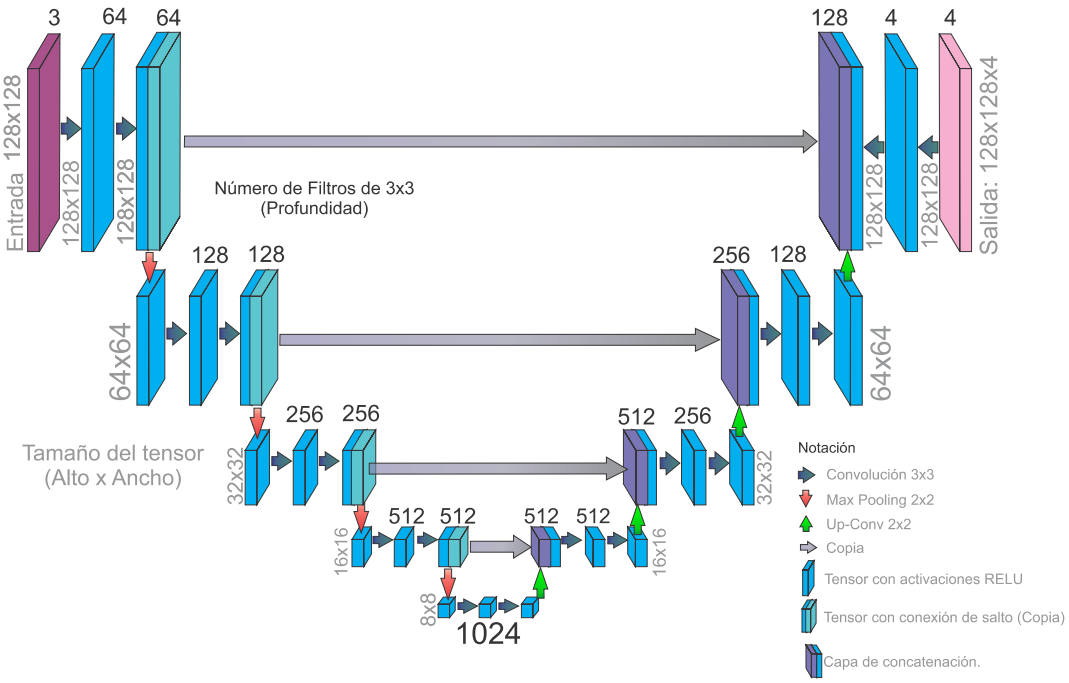


Figura 3. Arquitectura *U-Net*.

La arquitectura completa se muestra en la figura 3, en donde se aprecian las secciones de codificación (*encoder*) a la derecha.

2.5. ARQUITECTURA *DEEPLAB*

Múltiples estudios como [30, 34, 36], demuestran la alta efectividad que pueden llegar a tener configuraciones de redes profundas, basadas en arquitecturas de capas convolucionales de agrupación piramidal. Uno de los casos más exitosos ha sido la red conocida como DeepLab [30], en donde, utilizando una configuración conocida convolución dilatada, los algoritmos pueden realizar estimaciones con alta precisión, las cuales son invariantes a la escala e invariantes a la pose de los objetos en una escena.

Este tipo de algoritmos, ha sido utilizado en agricultura de precisión para el conteo de especies, e incluso la estimación de hojas o frutos que una planta puede tener [37, 38].

DeepLab es una red profunda cuya arquitectura (ver figura 6) inicia con un Encoder, que es una serie de capas piramidales convolucionales encargadas de la extracción de características, las cuales son invariantes a la escala, y termina con un Decoder, que permite, a partir de un tensor de características, dibujar una imagen con las etiquetas que cada píxel ha recibido, según las clases con las que fue entrenada la red. Esta red incorpora un algoritmo utilizado en el cálculo de las transformadas wavelet, denominado *algorithme à trous* que traduce del francés, algoritmo con agujeros o dilatado. También se conoce con el término *Atrous convolution*. La convolución dilatada agrega un salto *(stride)* a la forma en cómo se realiza una convolución normalmente. Su comportamiento se describe en la ecuación (3), en donde, para cada ubicación i en la salida y y en el filtro w, la convolución dilatada se aplica sobre el mapa de características de entrada x. El factor r permite determinar la dilatación espacial. Si $r = 1$ se tiene una convolución típica, mientras que con $r > 1$ se realizan saltos en el muestreo convolucional. La variable k representa el tamaño del *kernel*, Con el algoritmo *Atrous* se extraen características agregando variabilidad a los datos.

$$y[i] = \sum_{k=1}^{K} x[i + r \times k]w[k] \tag{3}$$

La primera versión de DeepLab presenta una red convolucional integrando configuraciones de redes profundas como VGG16 o ResNet101 como soporte principal. En estas arquitecturas se modifican algunas capas convolucionales para implementar el algoritmo *Atrous*, que realiza extracción densa de características. En el final de la red, se dispone un campo condicional aleatorio (CRF), entrenado junto con la red, y refina la segmentación de bordes y contornos de los objetos detectados. Debido a que la convolución dilatada entrega imágenes cuya resolución supera la de las imágenes ingresadas,

se efectúa una interpolación bilineal que reduce la resolución al mismo tamaño de la entrada y suaviza los bordes detectados de los objetos [39].

En la segunda versión la red DeepLab fue modificada adoptando el concepto de agrupación espacial piramidal, que se enfoca en la detección de objetos sin importar el lugar en el que el objeto se encuentre [30]. Como se muestra en la figura 4, una capa de agrupamiento espacial utiliza múltiples capas tipo *pooling*, cada una con un tamaño espacial diferente. El resultado de todos los *poolings* realizados se junta en un tensor que contiene la extracción densa de características. Otra de las ventajas de una pirámide espacial es que la imagen de entrada puede tener cualquier tamaño y resolución [40].

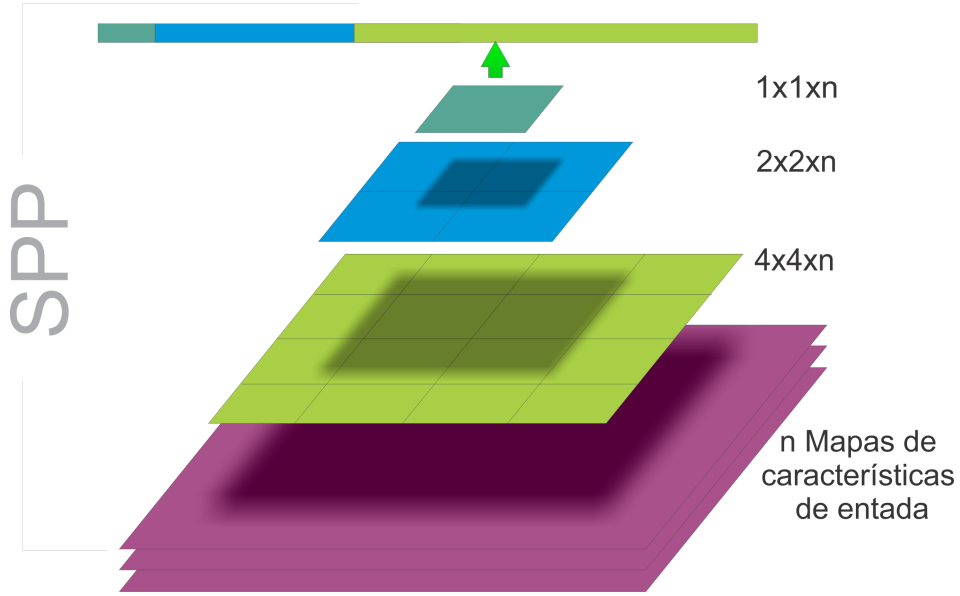


Figura 4. Capa de agrupamiento espacial piramidal [40].

En DeepLab V2, se junta el concepto de agrupamiento espacial piramidal y la convolución dilatada, con lo cual, se obtiene una capa de agrupamiento espacial piramidal dilatado (Llamada en inglés *Atrous Spacial Pyramid Pooling* o *ASPP*) como se aprecia en la figura 5, tal cambio mejora la detección de objetos en múltiples escalas y la precisión de la red [30].

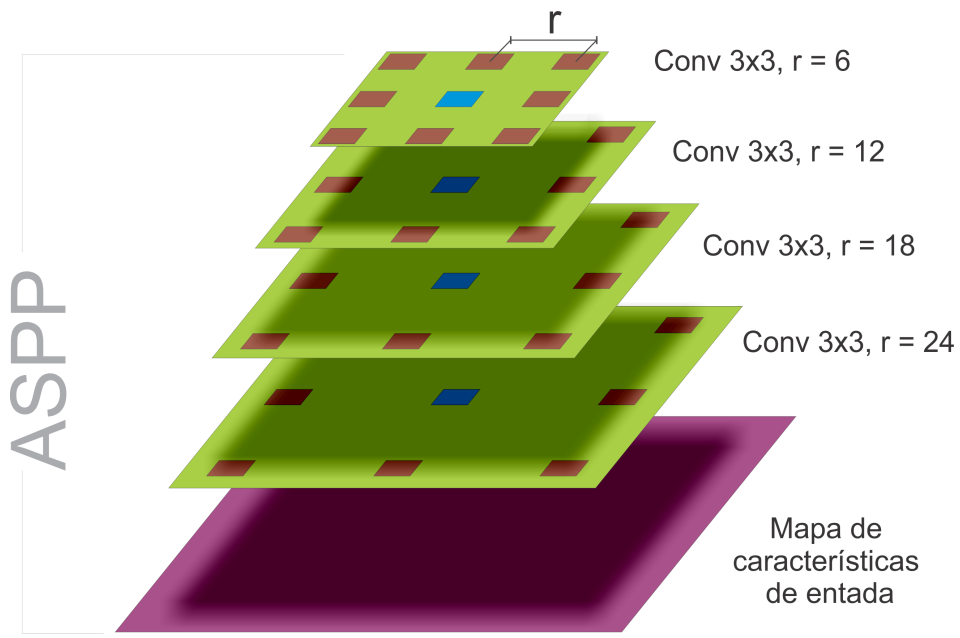


Figura 5. Capa de agrupamiento espacial piramidal con dilataciones *ASPP* [36].

En 2018 fue publicada la tercera versión de la red DeepLab, en la cual se replantea la forma en la que se realiza la convolución *Atrous*, superando los resultados obtenidos en sus versiones anteriores. Dentro de sus cambios más relevantes se encuentra la normalización del batch en las capas *ASPP* y la eliminación del CRF al final de la red. En DeepLab v3, se presenta una arquitectura con *ASPP* que reduce 8 veces la imagen respecto al tamaño de entrada, la cual posteriormente es aumentada mediante interpolación bilineal, para coincidir la salida con el tamaño original.

La última versión es DeepLab v3+, en la cual se cambia la interpolación lineal por una estructura convolucional que aumenta la resolución de la imagen en cada capa, es decir, se utiliza la red DeepLab v3 como codificador (encoder) y se utiliza una estructura convolucional como decodificador, tal como se aprecia en la figura 6. Tales modificaciones en la arquitectura, mejoran el desempeño de la segmentación semántica respecto a las arquitecturas predecesoras.

A pesar de su alto desempeño, este tipo de algoritmos requieren una capacidad considerable de cómputo, tanto en procesadores (CPU) como en unidades de procesamiento gráfico tales como GPU's para su entrenamiento.

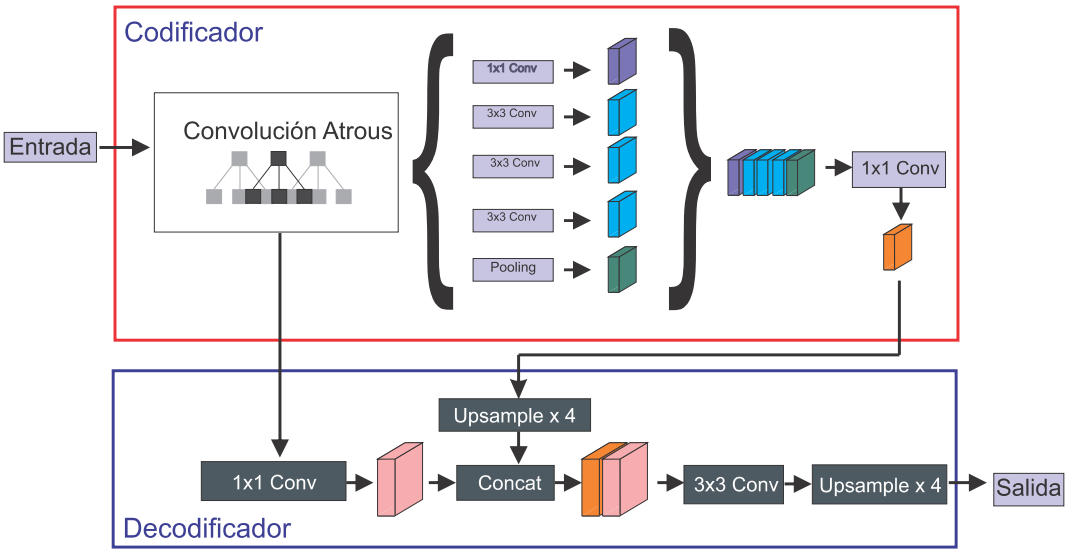


Figura 6. Arquitectura DeepLab v3 para segmentación semántica, [36].

La importancia del uso de segmentación semántica en especies vegetales radica en la necesidad de dar seguimiento en detalle a la irrigación de cada una de las plantas. Como se puede ver en [41, 42, 43] medir individualmente la irrigación permite conocer las cantidades óptimas de abono, riego y herbicidas que una planta necesita para que su crecimiento sea adecuado y la carga de frutos producida aumente o se mantenga. Otra ventaja es la optimización de recursos al evitar el desperdicio ya que tradicionalmente se toman algunas medidas y se aplican fertilizantes o herbicidas a todos los cultivos por la dificultad de medir una a una cada planta.

2.6. ESTIMACIÓN DEL ESTADO DE SALUD DE CULTIVOS

2.6.1. Índices de vegetación

La estimación del estado de salud de un cultivo puede hacerse utilizando técnicas de sensado remoto, como son las imágenes aéreas. Estas imágenes pueden provenir de satélites o de Drones. La principal característica que deben tener los lentes para el sensado remoto en aplicaciones de agricultura es que capturen los canales de color: rojo [570, 780] nm, verde [490, 570] nm, azul [400, 490] nm e infrarrojo cercano (NIR) [780, 1400] nm. Con la información reflejada en estas longitudes de onda se pueden calcular diferentes índices de vegetación. Tal es el caso del índice de vegetación diferencial normalizado o NDVI por sus siglas en inglés. Este índice es uno de los más utilizados en aplicaciones de agricultura de precisión [7, 24]. La razón de su amplio uso es porque las hojas verdes absorben la luz cuya longitud de onda está entre los 600 y 700 nm y reflejan las longitudes de onda de los 700 nm hasta los 1000 nm, es decir, si se aprecia más luminancia infrarroja que roja es por que existe capa vegetal en la imagen [19, 44]. Por consiguiente, la forma de calcularlo se aprecia en la ecuación (4), en donde el canal de salida resultante NDVI se calcula utilizando la radiación capturada por cámaras en el espectro de color rojo o R y el espectro infrarrojo o NIR.

$$NDVI = \frac{NIR - R}{NIR + R} \tag{4}$$

Como resultado, el canal NDVI tendrá valores en el rango entre -1 y +1. Los valores negativos corresponden a cuerpos de agua o nubes, mientras que los valores positivos cercanos a cero indican suelos o rocas. Así mismo, valores cercanos a 1 permiten evidenciar la presencia de plantas cuya actividad foto-sintética es alta.

Si bien, el NDVI permite hacer una estimación del estado de actividad foto-sintética de las plantas, este puede verse sesgado dependiendo de las especies que salgan en determinada fotografía. Es decir, si una foto se encuentra en presencia de especies con alta refracción de la luz infrarroja, aquellas que no tengan un índice de refracción tan alto aparentarán valores de plantas con falta de irrigación [45]. Para validar la situación descrita es posible realizar visitas en terreno, midiendo en las hojas de las plantas la firma espectral y cruzar la información con las imágenes. Otra alternativa es calcular el índice discriminando especies para poder determinar cuál es el estado de irrigación por plantas de la misma especie [46].

Una vez se adquieren imágenes multi-espectrales con Drones, es posible realizar una composición donde se visualice algún índice de vegetación y se aprecie el estado de irrigación de forma visual en una imagen. Esta información permite a los agricultores tomar las acciones necesarias para el cuidado de un sembradío [24].

2.6.2. Estimación del estado de salud con imágenes aéreas

Una de las premisas de la agricultura de precisión, es realizar el monitoreo constante del área cultivada [47]. Para esto existe una variedad de técnicas, como es el caso de las redes de sensores, que entregan grandes cantidades de información del suelo y de las plantas. Estas redes requieren de la instalación de nodos interconectados, así como suministro energético para cada uno de ellos, que puede llegar a ser una dificultad cuando se trabaja en grandes extensiones de tierra [5]. Otra de las técnicas, que ha tomado protagonismo gracias a la facilidad de despliegue y el bajo costo, es el uso de Drones para las tareas de sensado remoto y seguimiento de diferentes tipos de cultivos [19]. Mediante Drones con cámaras multi-espectrales, es posible realizar la estimación del estado de salud de las plantas en un cultivo, así como el análisis de algunos de los nutrientes que requieren y el estudio de zonas donde la irrigación no es adecuada [13, 48, 49]. La importancia de

tomar imágenes multi-espectrales a cultivos se debe a que esta técnica permite estimar el estado de irrigación de las plantas. Generalmente, en condiciones de suelo seco, las variedades vegetales sufren de estrés cuando el agua que evaporan debido a la actividad foto-sintética, es menor que el agua que pueden absorber del suelo. Tal situación lleva a las plantas a cerrar sus estomas y segregar hormonas que se encargan del cierre de las hojas para evaporar menos cantidad de agua, por consiguente, la fotosíntesis disminuye y la condición de sequía origina una menor producción de cosecha. Mediante el análisis del estado de irrigación, es posible prevenir que una planta se seque y aumentar la cantidad de cosecha que esta produce [18].

3. CAPTURA DE IMÁGENES AÉREAS MULTI-ESPECTRALES

En este capítulo se detallan los aspectos para la captura de imágenes con sensores multi-espectrales de bajo costo, así como los algoritmos planteados para la sincronización de múltiples cámaras utilizando la geo-localización y los sistemas de navegación de un Drone.

Para el desarrollo del proyecto, se cuenta con un drone DJI Phanthom 4 Pro, el cual, incluye una cámara que tiene un sensor de tecnología CMOS con resolución de 20Mpx. La cámara mencionada viene de fábrica con un filtro infrarrojo que permite la captura de imágenes en canales rojo, verde y azul, mientras rechaza la saturación proveniente de la radiación en el segmento de infrarrojo cercano o NIR. Esto es bueno para la captura de imágenes a color, ya que impide que el infrarrojo sature los canales rojo y verde, pero no permite registrar imágenes multi-espectrales. Por consiguiente, es necesario agregar un segundo sensor que venga de fábrica sin el filtro de infrarrojo cercano integrado. Una alternativa de bajo costo para adquirir el canal infrarrojo es el sensor de imágenes CMOS: Sony IMX219, el cual tiene una resolución de 8 Mpx y alcanza a captar radiaciones hasta los 880 nm como la que se muestra en la figura 7.

La arquitectura del drone no permite conectar sistemas de terceros a bordo, tampoco brinda acceso a las baterías para la alimentación de otros circuitos. Tal situación agrega la necesidad de incluir un sistema independiente de captura de imágenes NIR. Equipar el drone con un segundo sistema requiere solucionar con software los problemas asociados a la captura de imágenes desde dos dispositivos sin interconexión entre sí como se describe en este capítulo.

3.1. SISTEMA DE CAPTURA DE IMÁGENES NIR

Se implementa un sistema de captura de imágenes NIR, utilizando el sensor óptico descrito, conectado a una tarjeta Raspberry Pi ™. Adicionalmente se conecta un sensor de posición satelital de referencia Ublox NEO 6M capaz de medir coordenadas cinco veces por segundo, con un error medio de 2.5 m en la estimación de la posición. Para garantizar la precisión en la estimación de la posición, se programa un servicio de consulta del módulo GPS, el cual lee los mensajes en formato *NMEA*. Este formato brinda la información de la posición, así como la intensidad de la señal y el número de satélites enlazados. Experimentalmente se determina que para obtener la desviación media de posición de 2.5 m como indica la hoja de datos, el GPS debe estar enlazado a 10 o más satélites. Por este motivo en el software descrito se desarrolla un módulo encargado de verificar constantemente el número de satélites a los que el sensor de posición está enlazado y envía un mensaje de error cuando son menos de 10 ya que el error observado en tales casos supera los 50 m. Para la conectividad inalámbrica del dispositivo de captura NIR, se agrega un radio WiFi de alta ganancia, que permite acceder al computador a una distancia de hasta 30 m con línea de vista. También se adicionan baterías que dan una autonomía al sistema de captura de imágenes infrarrojas de 4 horas de uso continuo. Adicionalmente, para visualizar capturas durante el vuelo, se crea un servicio en Python que permite, a través de mensajería instantánea enviar comandos de ajuste a la cámara, como lo es el tiempo de exposición, la resolución de salida de las imágenes y la sensibilidad (ISO) del sensor. También se añade una utilidad para solicitar imágenes de muestra desde la aplicación de mensajería instantánea, este complemento ayuda a verificar que las configuraciones funcionen adecuadamente en el campo antes de iniciar una captura, sin tener que terminar una misión de vuelo y descargar las imágenes en un computador.

En el dispositivo se programa un servicio que monitorea la escritura de un archivo con el nombre *COORDENADAS.JSON* y contiene las coordenadas donde se tomarán las imágenes. El archivo se ingresa antes del vuelo a través del protocolo FTP. Una vez el servicio detecta la existencia de un nuevo listado de coordenadas procede a medir una vez por segundo la posición GPS y la compara con todas las del listado. Si alguna coordenada es similar a la posición actual, teniendo en cuenta un criterio de alineación basado en un modelo auto regresivo de media móvil, se toma una fotografía. Luego de la captura se incluye la información de geoposición en formato EXIF y se almacena para su posterior carga al sistema de gestión de imágenes multi-espectrales.

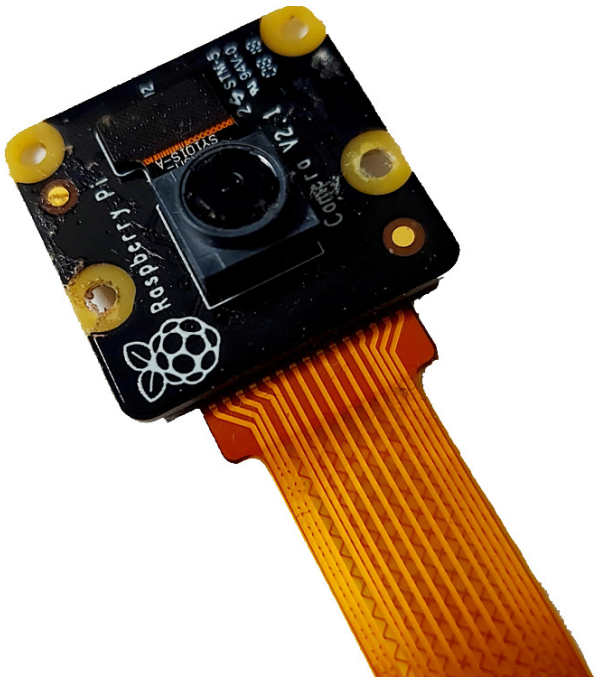


Figura 7. Cámara sin filtro infrarrojo utilizada.

3.2. ALINEACIÓN Y PRE-PROCESAMIENTO DE IMÁGENES NIR

El primer problema a resolver para la captura de imágenes multi-espectrales con múltiples lentes es la sincronía de la toma de datos, los cuales, deben estar espacialmente alineados para que la información medida sea similar. En general, los Drones Phantom de la marca DJI, no entregan ninguna alternativa para la captura de imágenes desde aplicaciones de terceros, ni de sistemas embebidos diferentes a su control remoto. Por tal motivo se equipa el sistema de captura de imágenes NIR, basado en la tarjeta Raspberry Pi™. El método para capturar fotografías en determinadas posiciones, consiste en calcular previamente las coordenadas de los centroides antes del vuelo. La información del lugar donde se debe registrar cada imagen, se almacena tanto en el drone (a través de su aplicación) como en el sistema de captura de canal NIR.

Con el fin de estimar la posición durante el desplazamiento del drone mientras se evita el ruido en la medición, se programa un algoritmo auto regresivo de media ajustable ($ARMA$) de factor 10. Es decir, se requieren de 10 muestras para la estimación de la posición media de un punto; por consiguiente a la medida actual se le pondera con el coeficiente 0.1, el resto de la ponderación, es decir el 0.9 restante, es para la media de coordenadas adquiridas previamente. La media se va desplazando con el ingreso de más muestras al modelo, tal como se describe en la ecuación (5).

$$LatLon = 0.9 \times LatLon_{previous} + 0.1 \times LatLon_{current} \tag{5}$$

Como segundo problema está la diferencia de resolución de los sensores ópticos. Esta dificultad es superada realizando un proceso conocido como *upsampling* en las imágenes con canal infrarrojo, mediante la aplicación de una interpolación bilineal de color.

Para encontrar el valor de color en un punto (x, y), conociendo el valor de cuatro puntos aledaños $P_{11} = (x_1, y_1)$, $P_{12} = (x_1, y_2)$, $P_{21} = (x_2, y_1)$, $P_{22} = (x_2, y_2)$, se describe la función de interpolación como:

$$f(x, y) = a_0 + a_1 x + a_2 y + a_3 xy$$

En donde los coeficientes de interpolación a_0, a_1, a_2 y a_3 se pueden encontrar mediante una regresión lineal, a partir de cuatro puntos correspondientes en las dos imágenes (x_1, y_1), (x_2, y_1), (x_1, y_2), (x_2, y_2) como se muestra en la ecuación (6).

$$\begin{bmatrix} f(P_{11}) \\ f(P_{12}) \\ f(P_{21}) \\ f(P_{22}) \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 & x_1 y_1 \\ 1 & x_1 & y_2 & x_1 y_2 \\ 1 & x_2 & y_1 & x_2 y_1 \\ 1 & x_2 & y_2 & x_2 y_2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} \tag{6}$$

Los puntos correspondientes en las dos imágenes se obtienen de forma automática con algoritmos que extraen puntos de interés como lo son *SIFT*, *SURF* y *ORB*.

Una vez se obtiene el canal NIR del tamaño de la imagen RGB se procede a su almacenamiento en el sistema de gestión de datos propuesto en este trabajo.

Como tercer problema, se tiene la respuesta de las cámaras en longitud de onda. La cámara sin filtro infrarrojo (NOIR), es sensible a la saturación del canal verde y rojo al estar expuesta a iluminación infrarroja. En el estado del arte, autores como [19] sugieren el uso de filtros de gel, como por ejemplo los filtros de la marca ROSCO denominados *Congo Blue 181*, *Deep Blue 120*, y *Dark Green 124*. Estos filtros dispuestos frente a la cámara evitan que las longitudes de onda equivalentes a los colores rojo y verde lleguen al sensor CMOS, no obstante, permiten el paso de la radiación infrarroja, la cual es

atrapada en el canal rojo. En síntesis, el canal NIR se puede extraer del canal rojo de la cámara NOIR con un filtro de gel azul en frente del lente.

En último lugar, la diferencia de distancia de los focos de las cámaras, ocasiona que la imagen vista en cada lente pertenezca a diferentes planos proyectivos. Es decir, si se sobreponen las imágenes de los diferentes lentes se verán los objetos con un desfase espacial. Este problema es ampliamente conocido en el estado del arte y puede solucionarse mediante la extracción de puntos clave que coinciden entre las dos imágenes, para conocer la homografía entre las mismas y poder remover la proyectividad.

Durante la remoción de la proyectividad es necesario eliminar la distorsión radial, mediante la calibración de las cámaras, tal como se describe en [20]. Para remover la perspectiva, se utilizan los algoritmos *SIFT* y *Harris* que permiten extraer múltiples puntos de interés entre dos imágenes con información similar. Debido al ruido, muchos de los puntos de interés pueden diferir entre una imagen y otra, con lo cual es necesario ejecutar el algoritmo *RANSAC* que permite excluir aquellos puntos que no tienen correspondencia entre las dos imágenes. Otras aproximaciones más actuales, adoptan técnicas de aprendizaje profundo, que han demostrado un mejor desempeño en tareas como registro de imágenes, en donde se hallan puntos clave y se sobreponen imágenes de un mismo lugar tomadas en diferentes momentos y desde diversas perspectivas *image registration*, pero no se abarca este enfoque ya que requiere entrenar un modelo a partir de una base de datos con imágenes y sus respectivas correspondencias marcadas por humanos [50].

Una vez se tienen los puntos de interés entre las imágenes, se realiza la estimación de la homografía y se transporta la imagen al espacio proyectivo deseado como se describe en [51].

Como salida del sistema de captura, se obtiene una imagen RGB proveniente de la cámara del drone, junto con una imagen de un solo canal que contiene la capa de radiación en el infrarrojo cercano, escalada y alineada espacialmente con la imagen RGB. Estas imágenes se envían a un sistema desarrollado de gestión para su consulta y almacenamiento adecuado.

4. ALMACENAMIENTO Y CONSULTA DE IMÁGENES

Una de las necesidades del proyecto es el almacenamiento y gestión de imágenes con más de tres capas de color. La metodología a emplear consiste en vincular las fotos RGB con su respectiva capa NIR a través de una base de datos relacional o de tipo SQL. Adicionalmente, como herramienta de trabajo es necesario un software para realizar tareas como: la consulta de los conjuntos de imágenes almacenados, el etiquetado de especies en una imagen como máscaras de segmentación semántica, la visualización de las diferentes capas de color guardadas, el cálculo de índices de vegetación, la construcción de imágenes panorámicas a partir de la base de datos y la predicción de máscaras en imágenes utilizando un modelo de Deep Learning. Se propone el desarrollo en lenguaje C# y Python, incorporando paquetes de licencia abierta (GPL) para el manejo de interfaces gráficas como es el caso de PyQt™. De igual forma se integran funcionalidades de la librería LabelMe [52], que es una herramienta gratuita, de código abierto, diseñada para la edición y visualización de máscaras semánticas en imágenes desarrollada en Python y C#.

4.1. MODELO DE BASE DE DATOS

La base de datos permite almacenar imágenes en un formato plano conocido como Blob. Este tipo de almacenamiento implica la decodificación de cada imagen en una matriz de números que representa la intensidad de cada píxel en 24 bits de color. Esta forma de almacenar datos no aprovecha las capacidades de compresión y hace necesario, en cada consulta, descargar un bloque de datos junto con la metadata para re-codificar los bytes en forma de imagen. Es decir, este tipo de almacenamiento consume más memoria y

toma más tiempo en realizar consultas de imágenes. Una alternativa que se plantea es guardar las imágenes (RGB y NIR) codificadas en formato JPEG en una carpeta de un servidor que puede consultarse utilizando el protocolo de transferencia de archivos FTP. De igual forma se dispone una carpeta accesible a través de una conexión FTP para el almacenamiento y consulta de máscaras de segmentación las cuales son imágenes en formato PNG.

En la base de datos SQL se crea una tabla que en sus campos guarda la siguiente información:

1. Vínculo FTP a la imagen RGB en formato JPEG.
2. Vínculo FTP a la imagen NIR en formato JPEG.
3. Vínculo FTP a la máscara de segmentación en formato PNG.
4. ID único para cada imagen registrada.
5. Coordenadas GPS en formato UTM.
6. Coordenadas GPS en formato WGS84 o magna-sirgas.
7. ID del conjunto de datos al que pertenece la imagen.
8. Fecha de escritura de la imagen.
9. ID de la curva de Hilbert de la imagen.
10. Vector de clases en la imagen.

Para la gestión de la base de datos, se define un módulo que permite hacer consultas SQL orientadas a insertar y eliminar registros. También se desarrollan funciones que permiten obtener los enlaces FTP a determinadas imágenes, según criterios de búsqueda

como son las clases, los índices de la curva de Hilbert, las coordenadas GPS, el ID del conjunto de datos y el identificador único.

4.2. CONSULTA RÁPIDA POR LOCALIDAD

El principio de funcionamiento de las búsquedas por criterios como máscaras semánticas o coordenadas se basa en una curva de llenado de espacio, en este caso la curva de llenado de espacio de Hilbert o (HSFC) [53]. Fue diseñada para mapear puntos ubicados en un espacio multidimensional a un vector de una sola dimensión con la ventaja de que conserva la localidad. Es decir, si dos puntos son aledaños en el espacio de múltiples dimensiones, también lo serán en el vector unidimensional.

Mapear puntos de un plano bidimensional ω (imagen) en una curva de orden α de Hilbert requiere de una transformación afín que puede ser lograda dividiendo el espacio ω en 2^{α} sectores cuadrados. Existen cuatro combinaciones que forman la secuencia de una HSFC, estas pueden ser representadas en forma de matriz equivalente como se puede observar en las ecuaciones (7) ,(8), (9), y (10) donde $(\beta, \beta 2)$ son la representación en forma de vector complejo de un punto en un cuadro unitario, y Q_n representa la partición cuadrada para el mapeo.

$$Q0 \begin{pmatrix} \beta 1 \\ \beta 2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \beta 1 \\ \beta 2 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{7}$$

$$Q1 \begin{pmatrix} \beta 1 \\ \beta 2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \beta 1 \\ \beta 2 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{8}$$

$$Q2 \begin{pmatrix} \beta 1 \\ \beta 2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \beta 1 \\ \beta 2 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{9}$$

$$Q3\begin{pmatrix}\beta 1\\ \beta 2\end{pmatrix} = \frac{1}{2}\begin{pmatrix}0 & -1\\ -1 & 0\end{pmatrix}\begin{pmatrix}\beta 1\\ \beta 2\end{pmatrix} + \frac{1}{2}\begin{pmatrix}2\\ 1\end{pmatrix} \tag{10}$$

Desde el punto de vista de la complejidad algorítmica, la búsqueda de un punto con coordenadas bidimensionales tiene una complejidad acotada por $O(n^2)$. Utilizando el mapeo a una curva de llenado de espacio, la complejidad se reduce a una cota superior de $O(log(n))$ [53]. Adicionalmente con esta técnica se puede crear un índice que permita ordenar las imágenes por similaridad en su contenido, como por ejemplo, tipo de cultivo que tienen o estado de irrigación.

El número de puntos del vector, depende del orden de la curva de Hilbert, es decir, a mayor orden, menor será el tamaño de la partición sobre la cual se realiza el mapeo. Esto permite alargar la longitud de la curva, como se muestra en la figura 8 y acomodar coordenadas en tantos puntos como se necesite. Para la selección del orden de la curva de Hilbert en determinado espacio de imágenes, se tiene en cuenta que la curva tenga, al menos, el doble de puntos que las coordenadas calculadas para un conjunto de imágenes determinado.

4.3. ALMACENAMIENO

Se define la funcionalidad de almacenamiento como un conjunto de algoritmos encargados de permitir que las imágenes multi-espectrales puedan ser añadidas y consultadas. Estos algoritmos vinculan cada imagen junto con su capa de infrarrojo en la tabla principal de la base de datos SQL. Los algoritmos se implementan en Python y su objetivo es facilitar tareas como: descargar las imágenes del drone, crear los registros en la tabla SQL, cargar las imágenes a las carpetas FTP, calcular los índices de la curva de Hilbert y enlazar las imágenes RGB con su correspondiente capa NIR.

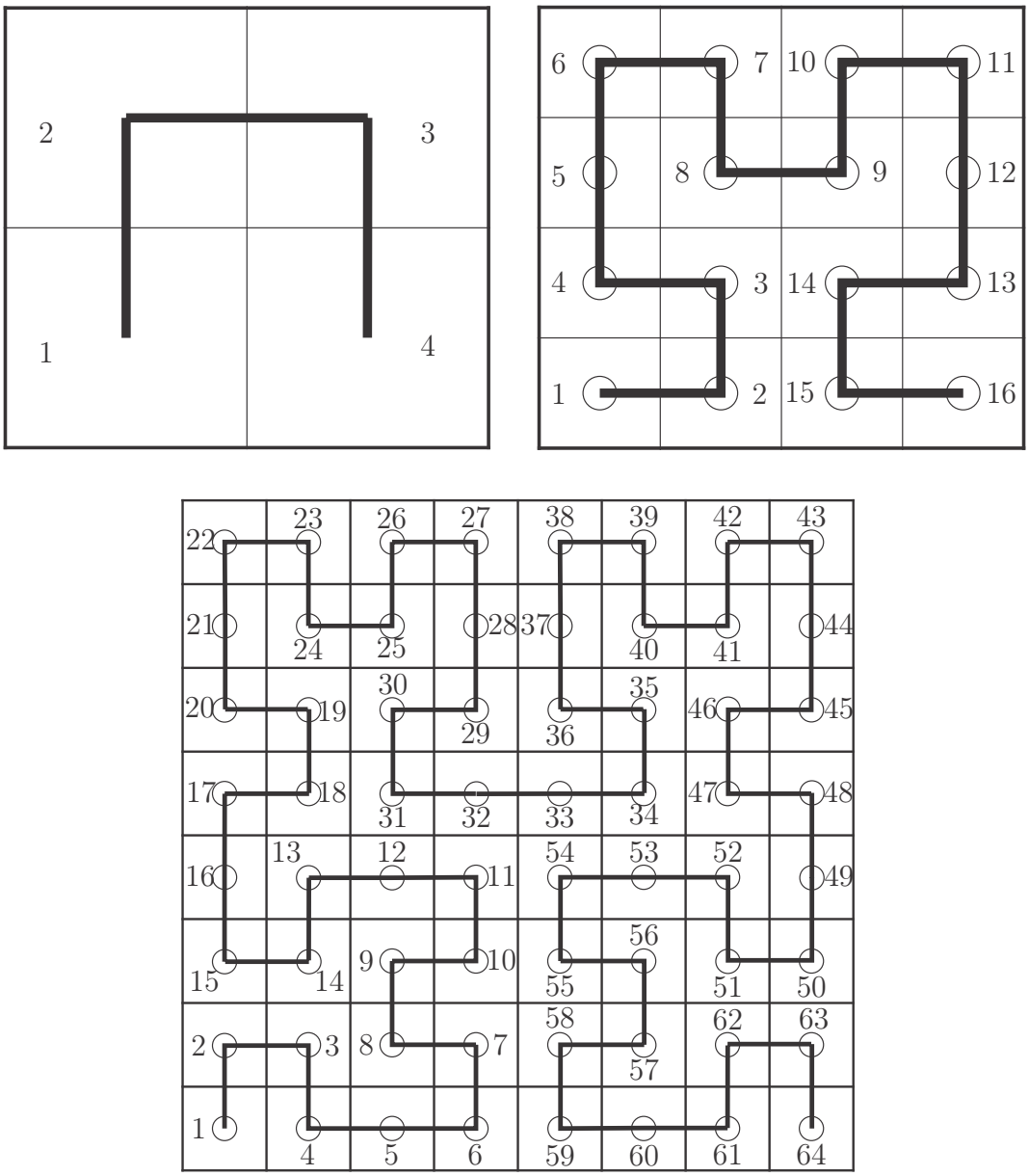


Figura 8. Curvas de Hilbert de orden 1, 2 y 3 [54].

La imagen de área amplia de un terreno puede estar compuesta por cientos e incluso por miles de capturas aéreas hechas por el drone. Además se pueden utilizar múltiples Drones para la captura de imágenes por sectores, por lo que los datos pueden sobrepasar la memoria de un computador convencional. La gestión del almacenamiento de grandes cantidades de imágenes se lleva a cabo con técnicas de almacenamiento distribuido

sobre un modelo SQL que provee confiabilidad y flexibilidad al sistema. Las imágenes se guardan en un sistema de tipo HDFS.

En la figura 9 se aprecia un esquema simplificado del modelo de almacenamiento, a partir de las capturas de un UAV en el campo, en donde, el drone captura dos imágenes con sistemas independientes a bordo, alineadas espacialmente. Una vez termina el vuelo, se comparan y ajustan los canales NIR con las imágenes RGB y por último se almacenan a través del sistema de gestión de imágenes desarrollado.

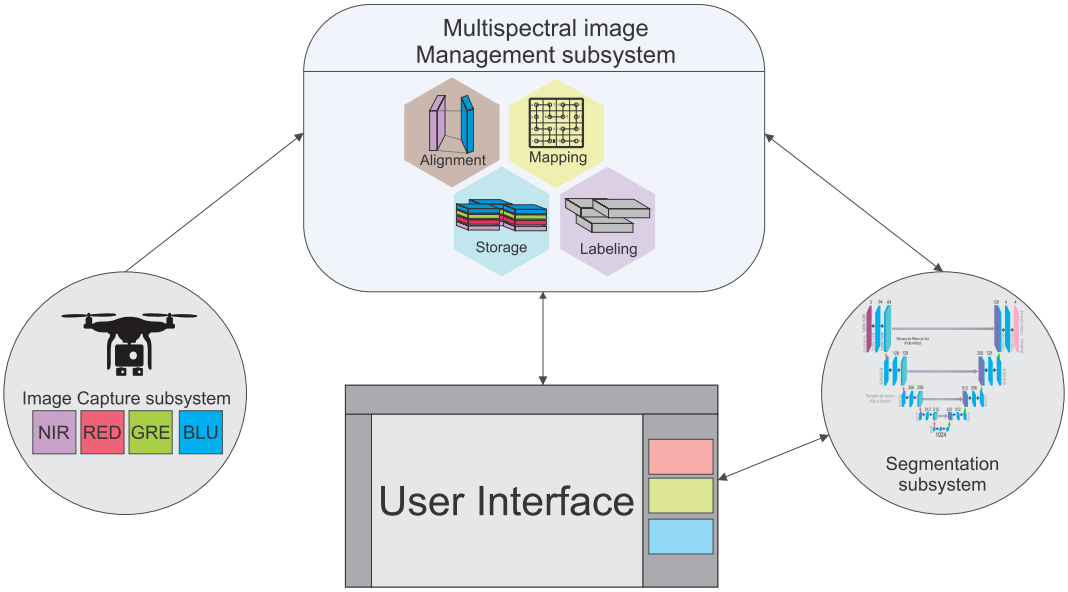


Figura 9. Esquema del sistema de captura y almacenamiento.

4.4. SISTEMA DE ETIQUETADO

En el estado del arte no se encuentran bases de datos anotadas que permitan realizar pruebas o modelos de segmentación semántica con cultivos Colombianos, en especial con cultivos como los de la región del Eje Cafetero, en donde se puede encontrar terrenos con café (*Coffea Arabica*) y plátano (*Musa Paradisciaca L*) juntos. Teniendo en cuenta que se necesitan crear modelos ajustados a los datos de la región, se plantea el diseño de

un componente de software para etiquetado basado en la alternativa de código abierto y uso libre LabelMe [52]. Este software permite generar archivos de etiquetas por especie en imágenes. Las etiquetas son almacenadas en ficheros de tipo JSON y se vinculan al ID de cada una de las imágenes en la base de datos. Adicionalmente se plantea el diseño de un programa que, a partir de las etiquetas en formato JSON, permita obtener las máscaras como imágenes en formato PNG, que son un requisito para el entrenamiento de la red profunda.

Para el manejo de las etiquetas, a través de comandos de línea se invoca al programa LabelMe, el cual, como se aprecia en la figura 10 permite dibujar polígonos sobre una imagen. Cada uno de los polígonos dibujados se puede asociar a un número entero positivo junto con una correspondencia a alguna clase definida por el usuario. Al finalizar la edición, el programa puede entregar etiquetas en formato de base de datos tipo Microsoft COCO [55], así como en formato XML con la estructura definida por la base de datos PASCAL VOC [56]. Se selecciona XML dado que permite guardar la etiqueta de una sola imagen, en contraste con el formato COCO (JSON) que almacena las etiquetas de todo el conjunto de datos en un único archivo. Por la forma en que se tratan los datos en la red profunda para el entrenamiento de modelos, se diseña un programa que convierte los ficheros de descripción de etiquetas en formato XML a una imagen en formato PNG. Las etiquetas por cada píxel serán valores enteros de 1 a k siendo k el número de clases a detectar. Se utiliza el componente de almacenamiento para la carga y vinculación a la base de datos de cada una de las etiquetas en formato PNG, obteniendo una base de datos anotada, que, con una consulta es capaz de entregar imágenes por zona y por coordenadas, junto con su respectiva máscara de segmentación.

Para anotar la base de datos con imágenes de la cámara Sequoia Sentera™ se utiliza una validación de una capa, en donde, a un usuario encargado de etiquetar cultivos se le presentan ejemplos de como debe hacerlo. Una vez el usuario termina la marcación

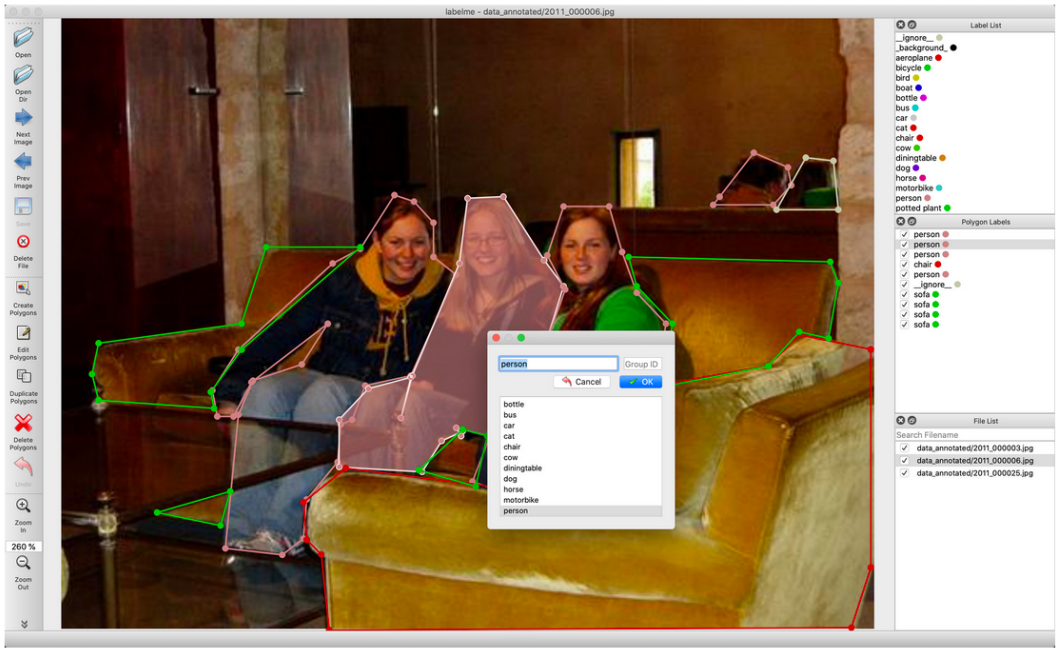


Figura 10. Interfaz gráfica de LabelMe [52].

de las imágenes asignadas, se pasan a revisión para que sean aprobadas como etiquetas semánticas por un usuario con el rol de validador. Esto es posible gracias a un campo en la base de datos, que relaciona las etiquetas con su estado, indicando si determinada imagen contiene o no etiquetas. En caso de contener alguna etiqueta se puede conocer si esta está revisada o si su etiqueta está pendiente de revisión.

En total fueron etiquetadas 270 imágenes, de las cuales se aprobaron 230 para el entrenamiento del modelo de aprendizaje profundo. Las imágenes restantes se descartan ya que no cumplen con la altura de captura entre 30 m y 50 m ni con el ángulo del lente perpendicular al plano del suelo.

Debido a que solo se contaba con un usuario etiquetador no fue posible calcular índices de concordancia entre etiquetadores. Sin embargo se pudo detectar que hace falta proveer al usuario de herramientas como tabletas de dibujo que le permitan señalar los polígonos con una mayor precisión de tal forma que en la etiqueta se diferencie el

límite entre el suelo y el cultivo. Se observó que muchas veces el etiquetador dibujaba rectángulos o secciones rectas que ocupaban gran parte de la imagen para facilitar su labor, abarcando en una etiqueta algunos píxeles del suelo. No obstante, las etiquetas se aproximaban a la realidad, siendo menores las intersecciones de la etiqueta con el suelo.

5. SEGMENTACIÓN DE CULTIVOS CON DEEP LEARNING

La segmentación semántica es el proceso de asignar a cada uno de los píxeles de una imagen una etiqueta de clase, es decir, determinar a que categoría de objetos pertenecen dentro de un listado predefinido [34].

Uno de los enfoques modernos para resolver tareas de segmentación es utilizando redes neuronales profundas o *deep learning* [38, 39]. La razón del auge de las redes profundas se debe a los resultados que ciertas arquitecturas presentan en el estado del arte segmentando diversos objetos en múltiples contextos [36, 10]. Para aplicar técnicas de *deep learning* en agricultura de precisión es necesario escoger una arquitectura adecuada y tener una base de datos debidamente anotada.

Para la selección de una arquitectura de red neuronal profunda se tiene en cuenta la firma de memoria del modelo, que, en casos como *DeepLab* [30] utiliza más de 130 millones de parámetros, requiriendo más de 3 Gb de VRAM para su ejecución y más de 11 Gb de VRAM para su entrenamiento. Si bien *DeepLab v3+* es la red que mejor desempeño presenta en el estado del arte para el momento del desarrollo del presente trabajo [36], su gran demanda de memoria VRAM dificulta el entrenamiento en equipos de cómputo convencionales. Es por esto que en publicaciones como [36, 30] los autores re-escalan las imágenes de entrada a tamaños pequeños y dividen el conjunto de datos en subconjuntos también conocidos como minibatches. Según el tamaño de las imágenes, el tamaño del minibatch y la base de la red, para entrenar DeepLab se requieren tarjetas de vídeo con capacidad entre 12 Gb hasta 64 Gb. Tales sistemas de cómputo requieren de mucha energía para ejecutar un algoritmo y son voluminosos y pesados, haciendo imposible ejecutar la inferencia desde un UAS. Por este motivo se utiliza una

implementación que tenga una menor firma de memoria. En específico la arquitectura de red *U-net* [34] que tiene 31 millones de parámetros, requiriendo para su operación una cantidad de 996 Mb de memoria de vídeo.

La selección de *U-net* permite estructurar la solución sobre un sistema embebido que puede implementarse a bordo de un UAS con el fin de segmentar las fotos adquiridas durante la captura (en línea). Se escoge como sistema de desarrollo, una tarjeta *Nvidia Jetson Nano* ™ que provee un computador integrado en una sola tarjeta, liviano, de consumo energético moderado, con 450 núcleos de GPU y con memoria VRAM de Gb. Este dispositivo puede ser configurado con una versión modificada del sistema operativo Ubuntu. Sobre el sistema operativo se instala un entorno conocido como *JetPack* que permite acceder a la VRAM y los núcleos de procesador gráfico. Este entorno provee acceso a una librería necesaria para procesamiento de alto desempeño en GPU llamada CUDA. Adicionalmente se instalan los entornos de trabajo para imágenes (OPENCV) y para aprendizaje de máquina (TensorFlow GPU).

Durante la propagación de una imagen de entrada a través de una red profunda, las operaciones convolucionales se efectúan con cierto número de filtros por capa y aumentan en cada etapa el número de volúmenes o dimensiones que un tensor de entrada tiene. El crecimiento de volúmenes representa un consumo elevado de memoria que debe restringirse para evitar saturar las capacidades de un sistema de cómputo. En el caso de la *Nvidia Jetson Nano* ™ se re-escalan las imágenes de entrada a un tamaño de 128 píxeles de alto por 128 píxeles de ancho. Respecto a los canales de color, se conservan tres de ellos que son: el canal infrarrojo cercano (NIR), el cual permite diferenciar especies vegetales de suelos y objetos inertes; el canal rojo (R) que contiene información de la absorción de luz de las hojas, y junto al canal NIR permite el cálculo del NDVI (como se muestra en la ecuación (4)) y el canal verde que también incluye información valiosa para la detección de plantas por color.

Estudios como [35, 57] muestran como se puede emplear *U-net* para aplicaciones de agricultura de precisión, con buenos resultados en estimación de cobertura terrestre, clasificación de sequía en maíz, y detección de capa vegetal. Incluso autores como [58] muestran como, modificando *U-net* es posible mejorar su rendimiento a cambio de un incremento en su firma de memoria.

Como flujo de trabajo para la segmentación de cultivos se suelen ejecutar tareas como:

1. Plan de vuelo y cálculo de centroides donde se realizarán las capturas sobre la zona de interés para cubrir el área con un traslape de imágenes del 80 % aproximadamente.

2. Captura y almacenamiento de imágenes multi-espectrales con un UAS sobre el terreno de interés.

3. Pre procesamiento de imágenes capturadas, el cual puede incluir filtros, alineación espacial, remoción de la proyectividad y escalamiento.

4. Inferencia de las especies detectadas a través del modelo de aprendizaje profundo previamente entrenado.

5. Recuperación de una máscara de segmentación a partir del tensor de salida del modelo profundo.

6. Re-escalado y filtrado de la etiqueta para que coincida con la imagen original.

7. Almacenamiento de la etiqueta de segmentación semántica.

6. ESTIMACIÓN DEL ESTADO DE IRRIGACIÓN

Uno de los aspectos más importantes en las actividades agrícolas es el riego [3] que permite llevar a cabo los procesos bioquímicos al interior de las plantas tales como el crecimiento celular, la fotosíntesis y la regulación de la temperatura a través de la transpiración. En primer lugar, el crecimiento vegetal se debe a dos procesos: la expansión celular y la división celular también conocida como mitosis. Ambos procesos requieren de la absorción de nutrientes desde el suelo, junto con agua para ser efectuados. Estudios como [59] muestran que una baja concentración de agua disminuye la actividad en las células con lo cual se evidencia que una planta crece hasta un 50 % menos en comparación con un espécimen correctamente irrigado.

En segundo lugar la fotosíntesis es el proceso por el cual la planta toma dióxido de carbono del aire, la radiación solar y electrones de moléculas de agua para generar glucosa que incorpora al metabolismo. La importancia del agua en la fotosíntesis se debe a que esta aporta los electrones necesarios para la reducción química del dióxido de carbono.

En tercer lugar, el agua es utilizada como medio de control de la temperatura superficial de las hojas en las plantas. Generalmente, los cultivos de frutas y verduras están a la intemperie, recibiendo la luz del sol. Debido a que la radiación solar es rica en componentes infrarrojos, se produce un efecto de calentamiento superficial que, de no ser controlado, puede generar un fenómeno conocido como estrés térmico. Ante tal fenómeno los estomas de las hojas tienden a aumentar la transpiración y a cerrarse para evitar la muerte de las células. Esto reduce la disipación de calor, por lo tanto minimiza el consumo de agua en la transpiración, pero al no refrescarse, aumenta la temperatura superficial afectando las hojas. Si la planta se encuentra en un déficit hídrico, ante el aumento de temperatura, las plantas presentarán cambios anatómicos y

morfológicos tales como la reducción del tamaño de las células, el cierre de los estomas y el cambio de la permeabilidad de las membranas celulares. Estos cambios derivan en un crecimiento reducido, poca productividad en los cultivos e incluso la muerte de plantas [60].

Es por lo descrito anteriormente que se debe mantener un buen estado de irrigación en las zonas donde se pretenden crecer cultivos, sin embargo, el exceso de agua en la tierra puede ocasionar impactos negativos como el aumento de costos de producción, el desplazamiento de los nutrientes del suelo y daños al medio ambiente por el uso industrializado del agua [61]. De forma tradicional la medición de irrigación en grandes extensiones de tierra se lleva a cabo utilizando de forma manual, incrementando los costos de producción como se detalla en [62]. Una de las alternativas para la estimación del estado de irrigación es mediante imágenes aéreas multi-espectrales. Estas se pueden adquirir desde satélites, con los que se pueden abarcar grandes zonas geográficas. Sin embargo, los satélites presentan una baja resolución a nivel de suelo, en estudios como [45] se aprecia que los satélites presentan resoluciones cercanas a los $20m^2$ por píxel. A este nivel, es posible estimar el estado de irrigación de forma general. No obstante existen impedimentos para el seguimiento de un cultivo con satélites, como las oclusiones generadas por las nubes, la falta de disponibilidad de datos en determinadas regiones y la baja frecuencia a la que se pueden obtener nuevas impresiones del suelo.

Para monitorear el estado de irrigación en terrenos amplios se pueden utilizar cámaras a bordo de Drones [19], esta técnica ayuda a estimar de forma ágil zonas con exceso o falta de humedad para su posterior control. La cantidad de agua en la capa vegetal puede ser determinada utilizando índices de vegetación como el caso de NDVI [24]. En la figura 11 se aprecia el estado de salud de un cultivo con base en diferentes valores de NDVI adquiridos a partir de la combinación de los canales infrarrojo cercano y rojo como se aprecia en la ecuación 4.

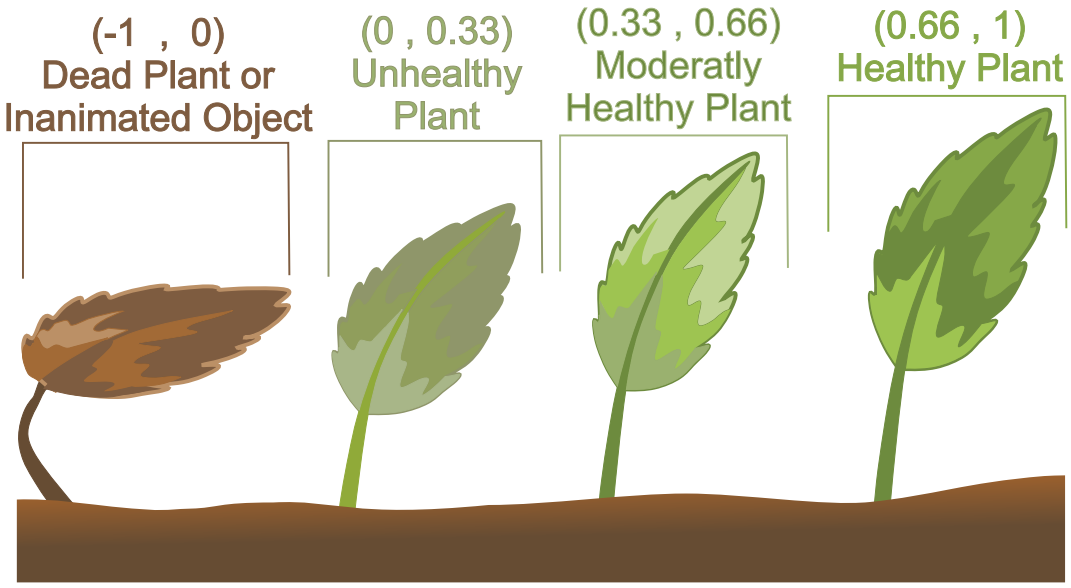


Figura 11. Significado de diferentes valores de NDVI en plantas.

Como no se cuenta con una cámara multi-espectral que permita medir el canal NIR directamente, se utiliza un enfoque como el tratado en [19] en el cual, combinando la información de una cámara sin filtro infrarrojo de bajo costo junto con una cámara RGB, se puede medir la cantidad de radiación infrarroja reflejada por las plantas.

La cámara de baja resolución con filtro de gel permite capturar el canal infrarrojo cercano. Este canal es escalado al tamaño de la imagen del drone con interpolación bilineal de color.

Una de las dificultades del cálculo del NDVI a lo largo de grandes extensiones cultivadas es el sesgo que se puede presentar al calcular el valor medio a lo largo de una imagen, si bien, el NDVI puede ser visualizado en escala de falso color, esto implicaría que un experto debe procesar y revisar una a una muchas imágenes multiespectrales. Para facilitar el trabajo de revisión de zonas de interés que puedan tener una baja irrigación se propone tomar las capas de color rojo (RED), verde (GRE) e infrarrojo cercano (NIR) a un modelo de aprendizaje profundo de máquina para segmentar los cultivos detectados. Sobre las máscaras de segmentación obtenidas se realiza el cálculo de los índices de

vegetación. De esta manera se conoce el estado de irrigación por cada especie vegetal en un cultivo. Esta técnica ayuda a minimizar el ruido de la medición y el escalamiento del canal NIR, también permite el monitoreo de forma rápida y la detección de zonas de bajo estado de irrigación de forma automática.

Adicionalmente, las imágenes adquiridas se pueden procesar de diferentes formas para poder estimar diversos parámetros de un cultivo. A cada procesamiento se le conoce como un índice de vegetación y es una representación esn forma de imagen del cálculo de las diferentes capas en el espectro de color adquirido que aportan información sobre los cultivos, los suelos, los cuerpos de agua y el calor reflejado en forma de radiación. Por tal motivo existen índices como son: NDVI, GVI, OSAVI, RVI, entre otros que ayudan a referenciar la vegetación teniendo en cuenta la reflectancia en el espectro infrarrojo. Según el índice la referencia se hace respecto al suelo o a la cobertura de la capa vegetal.

7. EXPERIMENTOS Y RESULTADOS

7.1. ADQUISICIÓN Y ALMACENAMIENTO DE IMÁGENES MULTI-ESPECTRALES

El montaje experimental para la adquisición de imágenes consta de una cámara sin filtro infrarrojo junto con una cámara RGB que viene de serie en un drone (DJI Phantom 4 pro). Estas cámaras se ubicaron en un montaje estático que permitía calcular la calibración estereoscópica con el fin de alinear las imágenes. El montaje estático garantiza que la matriz de calibración sea siempre la misma y que se puede corregir la proyectividad de las imágenes desde las diferentes cámaras.

En la figura 12 se aprecian las imágenes tomadas desde los dos lentes, el de color (RGB) del drone y el que no tiene filtro NIR, del sistema externo. Para efectos de alineación se hacen 30 capturas del patrón de calibración y se utiliza el método de Zhang [20]. Una vez se conocen las matrices de calibración es posible detectar la correspondencia entre la imagen del drone y la imagen de la cámara NIR, como se aprecia en la figura13. Con las correspondencias conocidas se procede a remover la proyectividad y ubicar la imagen NIR sobre la imagen RGB, el resultado se muestra en la figura 14.

Una de las dificultades resueltas con el trabajo de grado expuesto en [63] fue la sincronización de la captura de ambos lentes con un sensor GPS y el manejo de las coordenadas previamente calculadas. Se obtuvo un prototipo de sistema de captura de imágenes multi-espectrales a bordo de un drone, el cual permite adquirir imágenes con canales rojo, verde, azul e infrarrojo cercano. En los primeros vuelos de ajuste, se obtuvieron imágenes como las de la figura 15b, en donde se aprecia saturación de color en el canal infrarrojo debido a la intensidad de la luz solar. En la figura 15a se observa que los canales verde y azul, al no ser tan sensibles a la radiación infrarroja no presentan

saturación con la velocidad de captura seleccionada. Para evitarlo se reduce el tiempo de captura a 1/125 s que equivale al menor tiempo de apertura para capturar imágenes con el sensor Sony IMX.

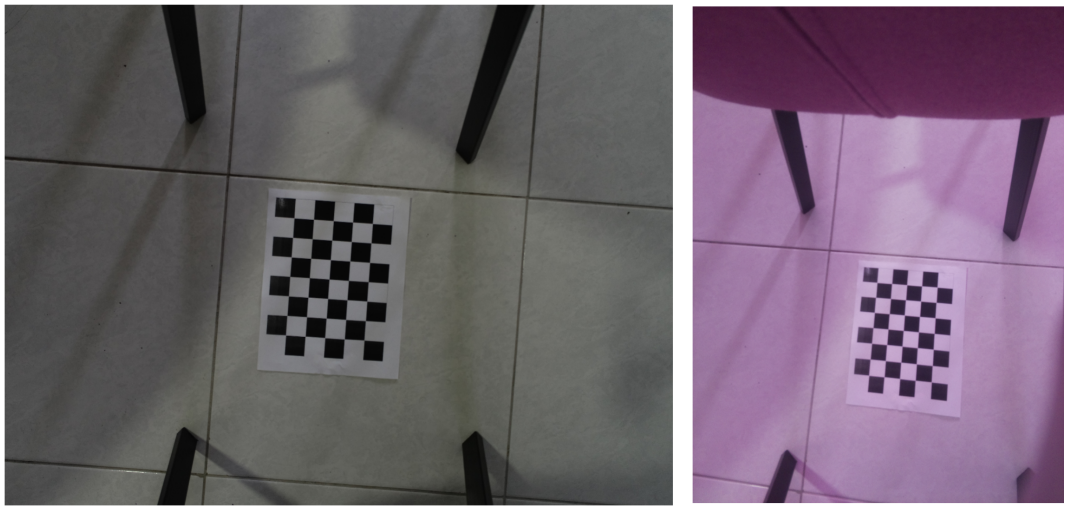


Figura 12. Captura desde las dos cámaras para calibración.

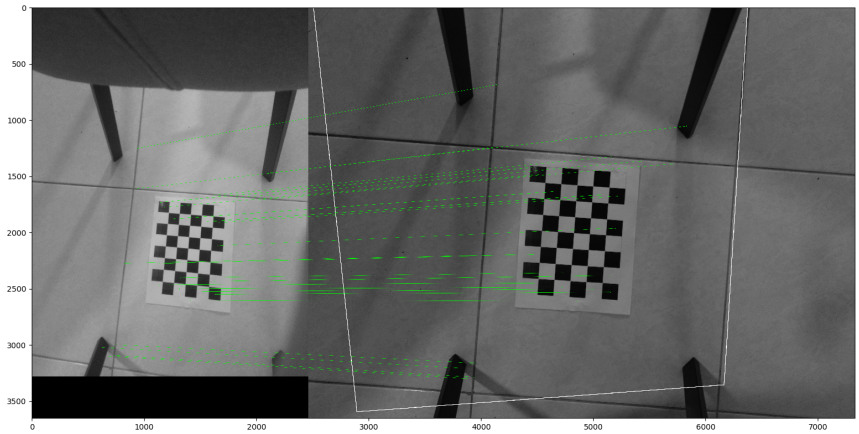


Figura 13. Detección de puntos clave para alineación.

La adquisición de nuevas fotografías para la calibración y validar los ajustes de iluminación y de tiempo de captura descritos tuvieron que ser suspendidos debido a la pandemia causada por la enfermedad del COVID19, que obligó a detener las actividades

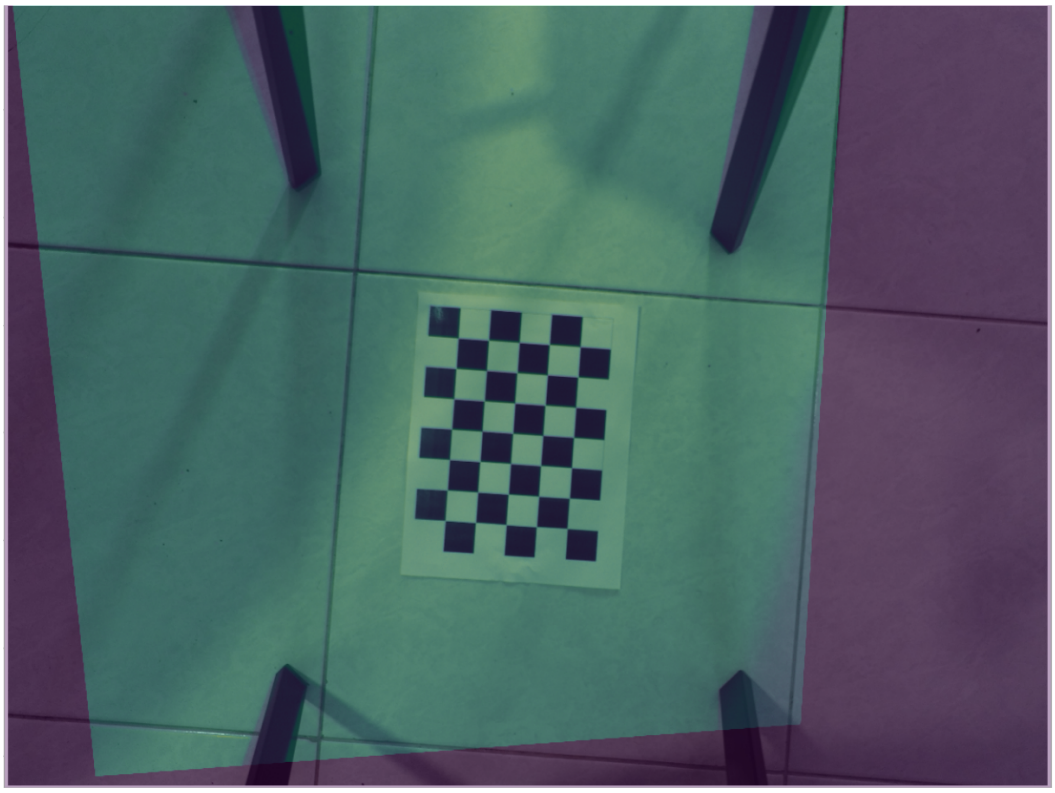
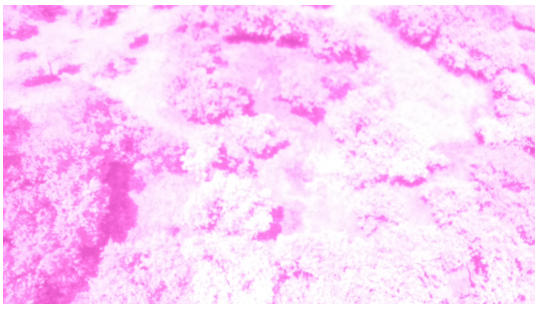
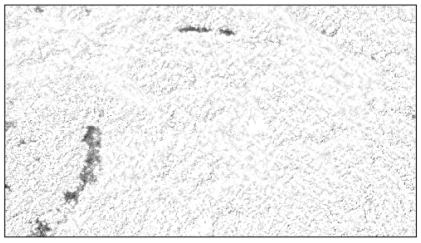


Figura 14. Imágenes alineadas.



(a) Imagen RG-NIR.



(b) Canal NIR saturado.

Figura 15. Imágenes capturadas con el montaje de cámaras experimental

en el terreno por más de 7 meses. Tal situación obligó a que se modificaran las actividades del objetivo específico 1, en el cual, ya no se empleó el conjunto de dos sistemas con lentes independientes y de diferentes resoluciónes montado en un UAS, sino que

se emplea una base de datos de 230 imágenes previamente adquiridas utilizando una cámara Sentera Sequoia ™. Respecto a la metodología planteada y el cumplimiento de los demás objetivos específicos, estos solo sufrieron retrasos en el tiempo mientras se tomó la decisión de modificar las actividades del objetivo específico 1.

La cámara Sentera Sequoia™ posee 5 lentes que capturan los canales rojo (R), verde (GRE), infrarrojo cercano (NIR) y frontera de rojo (REG). Si bien, las imágenes ya incluían la sincronización temporal de la adquisición, se necesitaba de la alineación fotogramétrica propuesta entre las capturas de cada uno de los cinco lentes para conformar las imágenes multi-espectrales necesarias. En síntesis se logró el cumplimiento del objetivo específico 1 con una cámara multi-espectral que integra cinco lentes en lugar de dos cámaras.

El conjunto de datos de 230 imágenes multi-espectrales fue adquirido en la sede "El Lembo" de las instalaciones del "SENA" ubicado a 20 km del municipio de Santa Rosa de Cabal en el departamento de Risaralda. Entre las imágenes se cuenta tres clases de cultivos que son: plátano (Musa Paradisiaca L), aguacate (Persea Americana) y café (Coffea Arabica).

La base de datos multi-espectral presenta una distribución de píxeles como se muestra en la figura 16, se evidencia un desbalance de clases, como en sucede en la mayoría de tareas de segmentación semántica, en las cuales la clase fondo es predominante.

Para organizar las fotos aéreas multi-espectrales adquiridas se utiliza el sistema de gestión diseñado. Este sistema cuenta con algoritmos encargados de almacenar los datos en un servidor y vincularlos utilizando una base de datos de tipo SQL.

La gestión es realizada en cuatro etapas, en la primera de ellas las capas de color son alineadas mediante la extracción de características con el algoritmo SIFT con el fin de corregir la proyectividad y las diferencias que existen entre los diferentes lentes. En se-

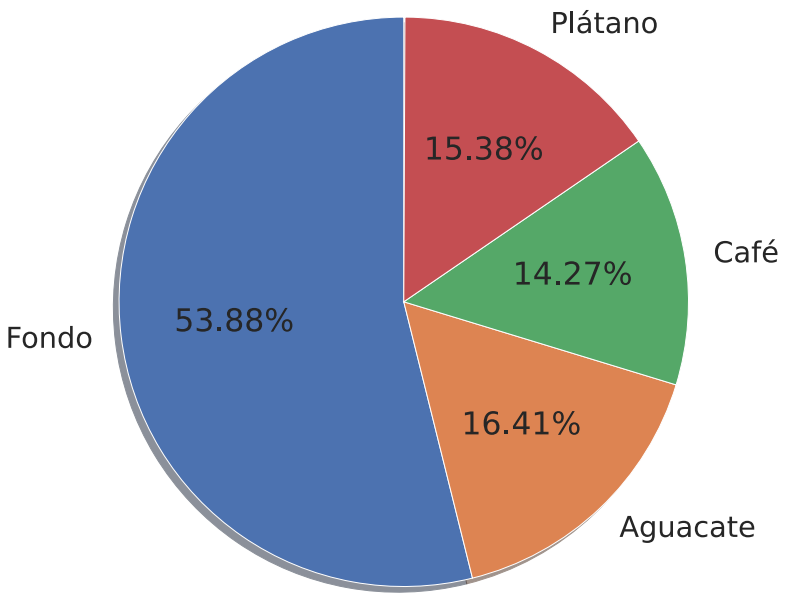


Figura 16. Distribución de las clases existentes en la base de datos.

gundo lugar se calcula la capa NDVI aplicando a las imágenes alineadas la ecuación (4). El resultado se almacena como una capa de color adicional. En el servidor se crea una carpeta por cada capa de color de las imágenes multi-espectrales con el fin de organizar el contenido. Estas carpetas son accesibles a través del protocolo de transferencia de datos FTP y sus vínculos se almacenan en una base de datos para la consulta de los componentes que conforman una imagen multi-espectral procesada.

En tercer lugar el algoritmo de carga extrae los meta datos como son la fecha y hora de captura, las coordenadas GPS y la información de la lente si está disponible, esta información es almacenada en una tabla de datos junto con el nombre de la imagen,

y las rutas de acceso FTP de cada una de las capas de color. Adicionalmente en este paso, se calcula el índice de la curva de Hilbert correspondiente a cada imagen para su organización por localización espacial.

Finalmente se cargan las imágenes a través de una conexión FTP en cada una de las carpetas dispuestas y se verifica que la carga haya sido exitosa para mantener el registro SQL en la base de datos. En caso de fallas en la carga el registro que relaciona las rutas y la información de GPS es eliminado para preservar la integridad de la base de datos.

Una vez las imágenes están en el sistema de gestión es posible realizar su etiquetado utilizando la herramienta basada en el software labelme [52]. En la figura 17 se presenta la ventana de trabajo del software, allí es posible seleccionar imágenes de una base de datos remota o imágenes multi-espectrales locales. Al escoger una imagen se visualizan por defecto las capas de color NIR, RED y GRE. Todas las componentes de la imagen multi-espectral junto con la capa NDVI calculada se pueden seleccionar en el panel del costado derecho, el cual se muestra en la figura 18.

Figura 17. Interfaz del software.

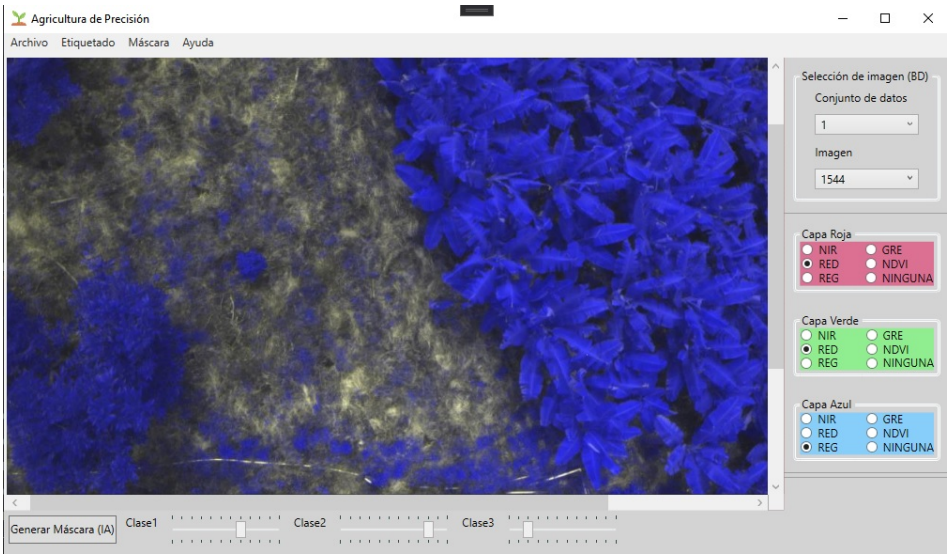


Figura 18. Muestra de la selección de diferentes capas de color.

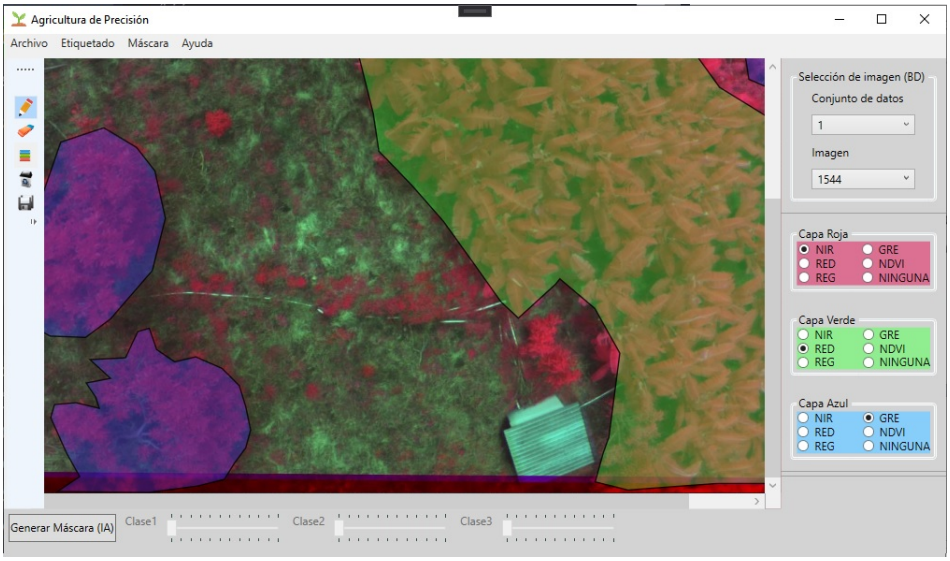


Figura 19. Modo de edición de etiquetas.

Desde el menú “Etiquetado” es posible habilitar la barra de herramientas que se aprecia en el costado izquierdo de la figura 19, también se puede visualizar algunas etiquetas creadas sobre la imagen. Las etiquetas se pueden dibujar como polígonos y se pueden guardar en formato PNG. Para validar la generación adecuada de etiquetas, estas se leen y se dibujan sobre una imagen. Debido a que las etiquetas son números enteros, se

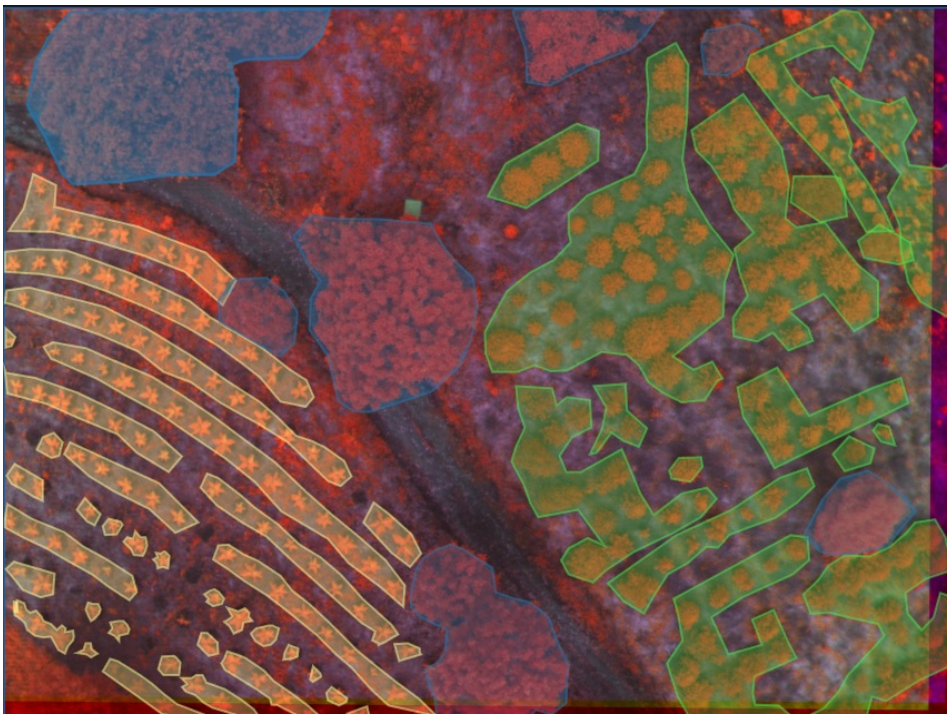


Figura 20. Etiquetas sobre la imagen NIR-RED-GREEN.

adiciona una correspondencia del número con una paleta de colores para que puedan ser diferenciables a la vista, como se puede observar en la figura 20.

Durante el proceso de carga de imágenes al sistema de gestión, las coordenadas GPS que registra la cámara son convertidas al sistema universal transversal de Mercator (UTM) y es calculado el índice correspondiente de la curva de Hilbert. El orden de la curva de Hilbert se escoge para que el número de puntos sea mayor al doble del número de imágenes capturadas para un conjunto de datos.

Como se aprecia en la figura 21, cuando se realiza una consulta por GPS y se buscan las imágenes aledañas, se obtienen aquellas que mejor se aproximan a la locación seleccionada sin importar si presentan rotaciones. Esta información permite construir imágenes panorámicas que cubren áreas amplias. Para la construcción se extraen y coinciden puntos clave entre las imágenes vecinas como se aprecia en la figura 22.

Figura 21. Ejemplo de consulta de índices de Hilbert aledaños.

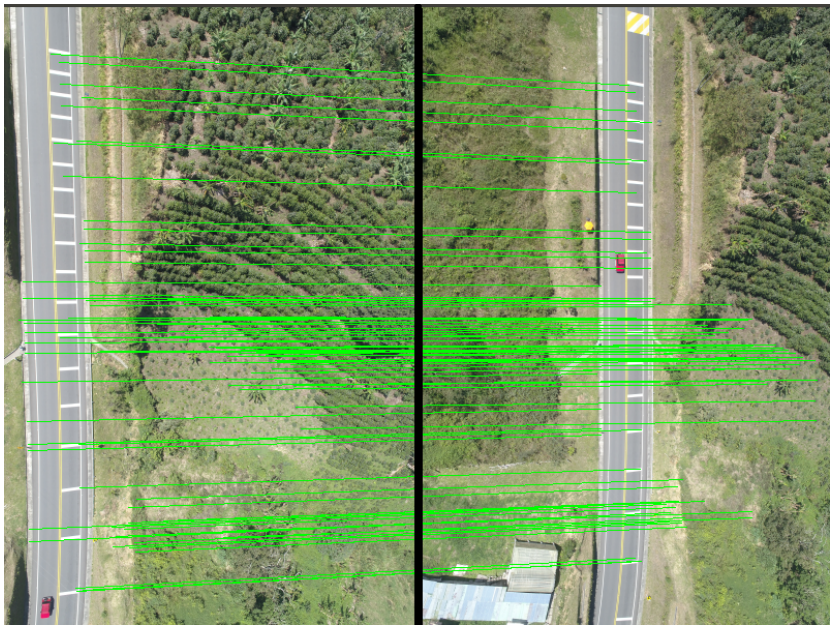


Figura 22. Coincidencias entre puntos clave.

Al repetir el proceso descrito se pueden obtener composiciones de imágenes como la que se observa en la figura 23 en donde se cubren aproximadamente $20.000m^2$ de tierra. Para corregir la distorsión en los canales de color se utiliza la calibración de cámara con lo cual se puede construir una imagen orto rectificada, esta se presenta en la figura 24.

Figura 23. Imagen construida con múltiples capturas.

Figura 24. Imagen construida teniendo en cuenta la calibración de cámara.

7.2. DETECCIÓN DE PLANTAS Y CÁLCULO DEL ESTADO DE IRRIGACIÓN

7.2.1. MONTAJE EXPERIMENTAL

El entrenamiento del modelo es realizado en la plataforma Colaboratory de Google, la cual permite disponer por lapsos de 12 horas de computadores equipados con tarjetas de video Nvidia Tesla k80. Este entorno ofrece de forma gratuita a desarrolladores la capacidad de ejecutar algoritmos en GPU o incluso TPU. Para el caso específico de las

GPU, cualquier usuario puede disponer de entre 8 a 12 Gb de memoria para operaciones (cantidad de memoria variable, dependiendo de la disponibilidad de los servidores de Google) que se pueden utilizar a través de un framework que corre instrucciones de Python.

Para realizar la inferencia se implementa la red *U-Net* ajustada para recibir imágenes de $128x128x3$. Las tres capas de color que componen la imagen son: infrarrojo cercano, rojo y verde. Esta selección se debe a que en estas capas es donde se aprecia en detalle el estado de irrigación y es más notoria la diferencia entre suelo y plantas. Se considera que el uso de las capas de color azul, límite de rojos y la capa NDVI pueden aportar en la detección, sin embargo, el modelo se diseña para ser ejecutado en una tarjeta NVidia Jetson™ Nano, la cual es un computador que puede implementarse a bordo de un drone y cuenta con 4 Gb de memoria RAM y un procesador con 452 núcleos GPU para el trabajo en paralelo. Esto permite realizar la inferencia a bordo desde el Drone para cargar al sistema las etiquetas detectadas.

El entrenamiento de la red se diseña dividiendo el conjunto de 270 imágenes en dos subconjuntos, el primero de ellos que cuenta con 189 de las imágenes (70 %) y el segundo para validación con 91 imágenes (30 %). El conjunto de datos de entrenamiento es generado desde el sistema de gestión de imágenes de área amplia, el cual entrega dos carpetas. Una de ellas, contiene las imágenes de capas R-G-NIR, la otra carpeta contiene las máscaras de segmentación semántica correspondientes a cada una de las imágenes. Cada una de las máscaras tiene el mismo nombre de las imagen a la que corresponde con el fin de facilitar su relación por fuera del sistema de gestión de datos.

Una vez descargadas las imágenes, se utiliza un algoritmo que aloja las fotos en memoria, junto con sus etiquetas en conjuntos o *batches* de 2 imágenes cada uno. La carga se realiza ubicando las correspondencias imagen-etiqueta en un contenedor de tipo *tensor* de TensorFlow. Adicionalmente se diseñan programas que utilizando la clase

tf.data.dataset que incorpora los algoritmos necesarios para el mapeo de etiquetas a nivel de píxel para el entrenamiento de la red de segmentación semántica.

Para calcular el gradiente se emplea la función de costo definida en (11) que es conocida como *Sparse Categorical Cross-entropy*. Esta función permite comparar la etiqueta predicha por la red con la etiqueta verdadera y calcula la perdida para múltiples clases. Así mismo, puede ser empleada cuando no existe una codificación de tipo *one hot* sino que hay etiquetas con valores enteros, como es el caso de la segmentación semántica. La función de costo itera sobre las K posibles clases predichas por la red profunda durante la propagación hacia adelante o *forward propagation*. Por cada clase toma el logaritmo natural de la etiqueta predicha multiplicado por cada píxel de la etiqueta original.

$$J(w) = -\frac{1}{K}\sum_{i=1}^{K} y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i) \tag{11}$$

Como algoritmo para la minimización de la función de costo (11) se emplea una versión modificada de *Adaptative moment estimation* o *ADAM*. Los autores de la modificación lo llaman Radam que proviene de las palabras *Rectified ADAM* [64]. Este optimizador es una extensión del algoritmo de gradiente descendente estocástico (SGD) ampliamente utilizado para entrenar modelos de aprendizaje de máquina. El algoritmo Radam define una tasa de aprendizaje independiente para cada uno de los parámetros de la red. Durante el entrenamiento los parámetros se adaptan según la velocidad media de cambio de las magnitudes de los gradientes. Esto permite acelerar la velocidad de convergencia de un modelo al variar la caída de la tasa de aprendizaje por década. Adicionalmente el algoritmo Radam incorpora una heurística denominada *warmup* o calentamiento. En el manuscrito [64] los autores demuestran como es posible mejorar el entrenamiento con la técnica descrita ya que se evita el sesgo de la red profunda a la distribución de probabilidad de las primeras muestras.

Para evaluar el desempeño del modelo se incluyen dos métricas, la primera de ellas la precisión (*Accuracy*) que indica cuantos píxeles quedan bien etiquetados en la predicción. La segunda métrica que se incluye sirve para estimar la similaridad entre una etiqueta y su predicción. Corresponde a una variante del índice Jaccard tambien conocido como intersección sobre la unión (IoU). Esta métrica calcula la media de intersección sobre la unión para cuando se detectan K clases siendo $k > 1$ y es denominada Mean IoU o MIoU. Se codifica una clase que permite, en cada época de entrenamiento almacenar el modelo con el mejor desempeño y el resultado de las métricas. Por último se configura el optimizador para dar un paro anticipado si tras 50 épocas no se evidencia una reducción en la función de costo, es decir, el algoritmo ha llegado a un mínimo pronunciado o tiene un sesgo considerable que le impide mejorar.

Para complementar la base de datos, antes del entrenamiento se aplican algunas transformaciones en las imágenes, como son rotaciones y recortes aleatorios según recomiendan los autores en [34], con el propósito de mejorar los resultados en presencia de bases de datos pequeñas. En este caso se duplicó tamaño del conjunto de datos inicial, esta técnica es mencionada como *Data Augmentation* y aporta cuando se pretenden obtener buenos resultados en una red profunda a partir de pocos datos [65]. Se omiten las transformaciones de perspectiva y de color ya que se considera que no contribuyen en la mejora de la detección. La distribución de la radiación en los canales de color que captura la cámara multi-espectral no es sensible a cambios de color o a desbalance de colores en canales como el rojo o el verde. Adicionalmente, el uso de imágenes aéreas tomadas entre 30 m y 50 m de altura sobre los cultivos con la cámara perpendicular al plano del suelo evitan que el modelo tenga que predecir plantas con transformaciones afines.

El entrenamiento se efectúa treinta veces, teniendo como criterio de parada el primer evento que suceda entre la ejecución de 200 épocas de entrenamiento, y el transcurso de

50 épocas sin que el modelo presente una mejoría en las métricas (*early stopping*). Los diez primeros experimentos se efectuaron utilizando como algoritmo optimizador *Adam*, en el cual se tiene una media de entrenamiento de 25 horas. Sin embargo, ninguno de los modelos obtenidos logró superar un 35 % en la métrica MIoU. Se evidenció una tendencia del modelo a sesgarse a los primeros ejemplos de entrenamiento. Respecto a los 20 experimentos restantes, fueron efectuados con el optimizador del estado del arte *Radam*. En este caso se evidencia una reducción notoria del tiempo de entrenamiento, así como una convergencia rápida. Se evidenció que los algoritmos presentaban un paro anticipado en una media de 2.5 horas. En la figura 25 se visualiza la precisión por épocas de entrenamiento, en color azul la medida sobre el conjunto de datos de entrenamiento muestra una tendencia creciente y no se evidencia un sobre entrenamiento del modelo. Respecto a los datos de validación se verifica que los resultados presentan un comportamiento que tiende a la mejoría, sin embargo, es posible mejorar el modelo utilizando más imágenes a la entrada del modelo.

Para validar el desempeño del modelo *U-net* entrenado se calcula la matriz de confusión como se presenta en la figura 2, en donde la clase 0 corresponde al fondo, la clase 1 corresponde a *Persea Americana*, la clase 2 corresponde a *Coffea arabica* y la clase 3 corresponde a *Musa Paradisiaca L.* Se evidencia que el modelo tiene un buen desempeño en la detección de las múltiples clases en el conjunto de datos. Adicionalmente se realiza el cálculo de cuatro métricas que son:

- Precisión (Accuracy): Indica la cantidad de verdaderos positivos sobre la totalidad de píxeles en la base de datos de prueba.

- Exactitud (*Precision*): Indica la cantidad de aciertos del modelo sobre el total de valores predichos como positivo.

- Sensibilidad (Recall): Muestra la cantidad de aciertos que son verdaderos positivos en relación con el total de predicciones.

- Puntaje F1 (F1 Score): Esta métrica indica el balance entre la precisión y la sensibilidad del modelo, permite revisar el desempeño del modelo cuando hay una distribución desigual de clases en los datos.

En la tabla 3 se observa que el modelo tiene una buena precisión y sensibilidad, es decir detecta de forma adecuada cada una de las cuatro clases presentes en los datos.

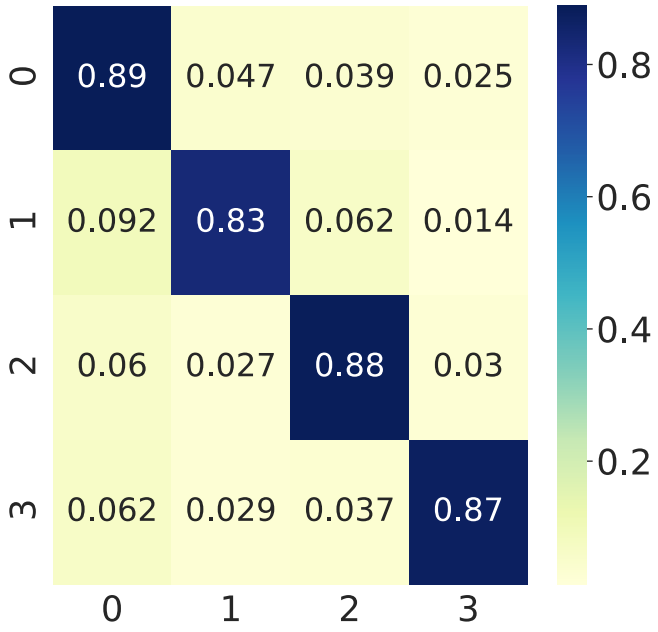


Tabla 2. Matriz de confusión del modelo.

En la figura 26 se aprecia la pérdida del modelo (*loss*) evaluada por cada época durante el entrenamiento. Con esta información se valida que el modelo realiza un ajuste de la función de costo constante sin que exista un sobre entrenamiento, pero se aprecia que requiere de un mayor número de imágenes para lograr mejores resultados. La forma de

Métrica	Clase 0	Clase 1	Clase 2	Clase 3	Media	Media ponderada
Precision	0.94	0.83	0.79	0.89	0.86	0.90
Recall	0.91	0.82	0.89	0.90	0.88	0.89
F1 Score	0.93	0.82	0.83	0.90	0.87	0.89
Accuracy					0.89	-

Tabla 3. Métricas del modelo

visualizar qué tan bien están ubicadas las etiquetas calculadas por la red, respecto a las etiquetas predichas es mediante la métrica de media de intersección sobre unión. Esta se puede apreciar en la figura 27 En donde se aprecia que durante la validación hay una tendencia al crecimiento, es decir, el modelo tiene la capacidad de generalizar las etiquetas sobre imágenes que no se le habían presentado previamente en la etapa del entrenamiento.

En la etapa de etiquetado, muchos de los polígonos dibujados sobre las plantas abarcan píxeles que pertenecen al suelo. Esto se debe a que el etiquetador humano disponía de un ratón de computador para dibujar las etiquetas y con esta metodología muchos de los píxeles que pertenecen al suelo son marcados como parte de las plantas. En la imagen 28 se puede apreciar en el medio la imagen con la etiqueta dibujada por humanos, la cual abarca píxeles dentro de la etiqueta “café” que en realidad pertenecen al suelo. No obstante, en la imagen del costado derecho, donde se muestra la máscara predicha por la red profunda se puede apreciar que el algoritmo es capaz de discriminar los píxeles que pertenecen al suelo respecto a los que pertenecen a la clase çafé". Es decir, el sistema puede funcionar como un soporte para el ajuste fino de las etiquetas realizadas por humanos.

Una consideración de los buenos resultados se debe a la distribución medianamente balanceada de clases, debido a que en las imágenes capturadas no existe un des balance de clases extremo como pasa en [39], en donde la clase fondo abarca más del 80 % de los píxeles en el conjunto de datos. Por lo tanto no es necesario aplicar funciones de costo

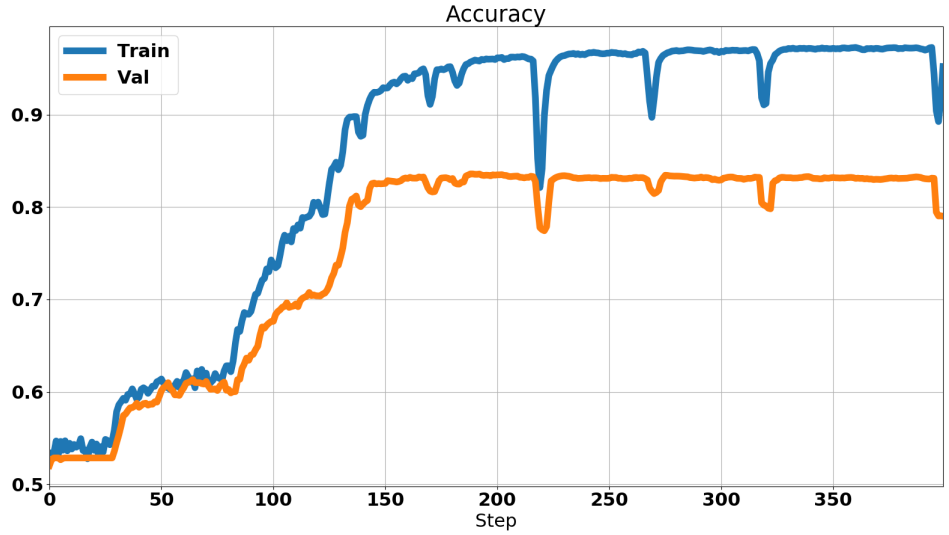


Figura 25. Precisión del modelo.

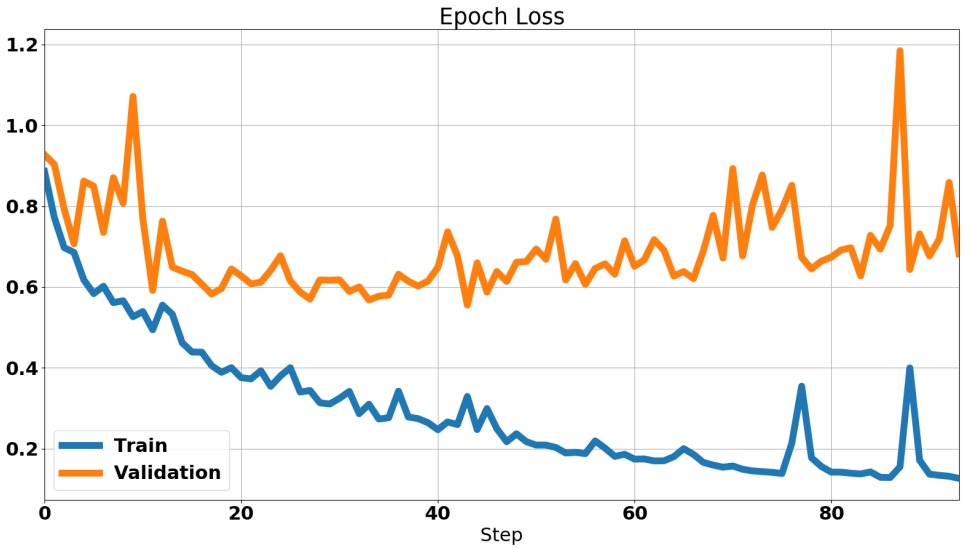


Figura 26. Pérdida evaluada por épocas.

que se encarguen de alterar la ponderación de la clase fondo, de tal forma que eviten que la red se entrene para aprenderse el fondo y no el objeto de interés.

En la figura 28 se aprecia que el sistema es capaz de detectar y discriminar adecuadamente píxeles de suelo que no pertenecen al cultivo, incluso si la etiqueta hecha por humanos no tiene un buen ajuste a nivel de píxel. Como trabajo futuro se propone

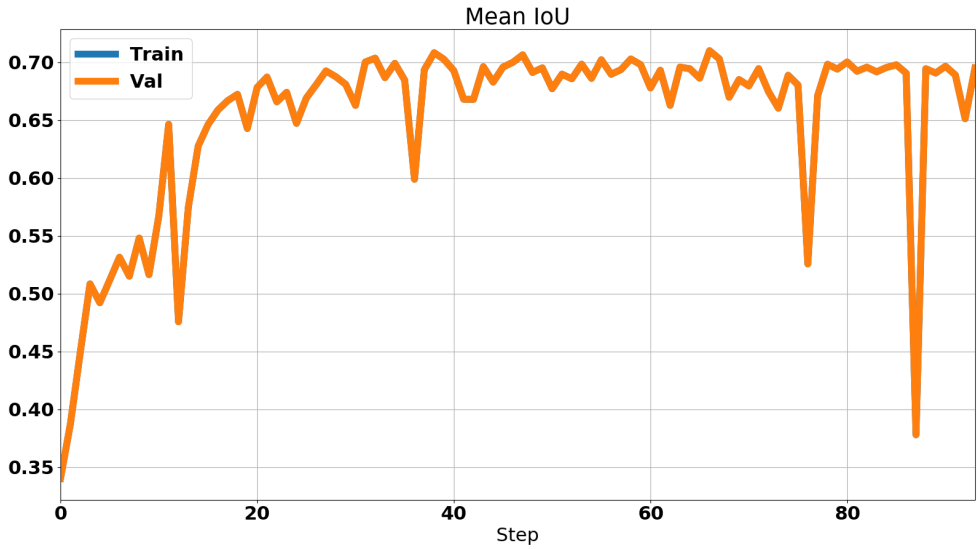


Figura 27. Media de intersección sobre unión.

utilizar el sistema como una opción para el etiquetado asistido o para el etiquetado bajo una metodología *human in the loop* que ayude en la creación de bases de datos de cultivos.

Figura 28. Resultados de la arquitectura entrenada.

Algunas muestras del algoritmo de segmentación semántica son presentadas en la figura 29; en el costado izquierdo se muestra la imagen de un cultivo construida con las capas rojo, infrarrojo cercano y verde. En el centro están las etiquetas creadas y validadas y en el costado derecho las predicciones del modelo *U-net*.

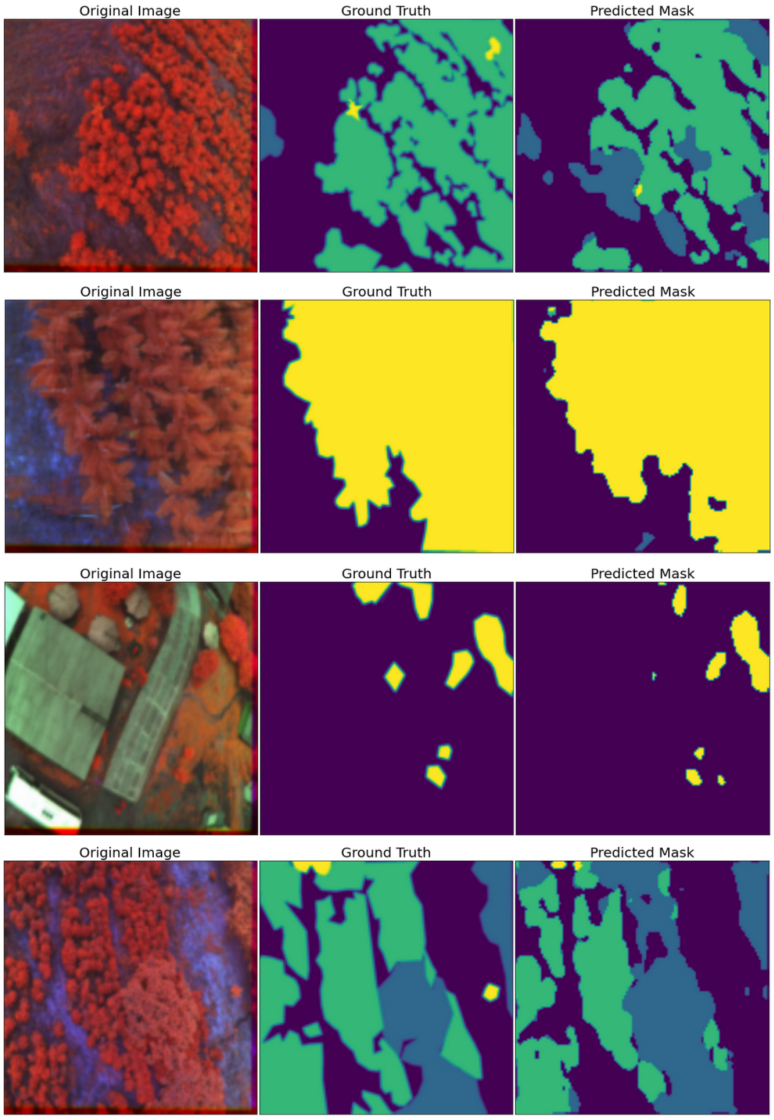


Figura 29. Resultados de segmentación de algunas imágenes de validación.

Tal como se aprecia en la figura 30 se utilizan las máscaras correspondientes a la clase Aguacate, para extraer los árboles detectados con la red profunda y así calcular el estado de irrigación ajustado solo a los píxeles recortados.

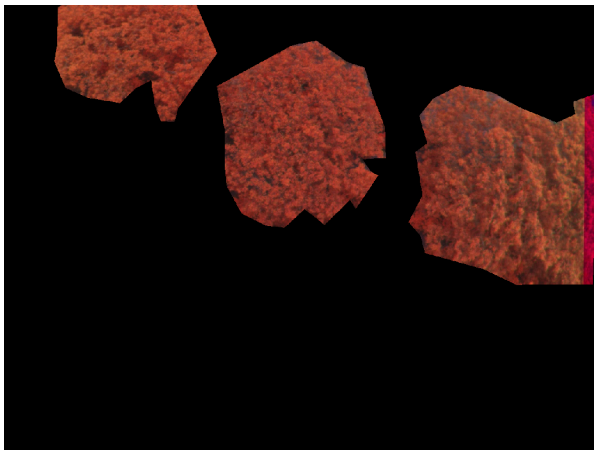


Figura 30. Segmentación de Aguacate.

Con el mismo procedimiento se segmenta y evalúa el estado de irrigación para píxeles que representan cultivos de café, esto se muestra en la figura 31. Finalmente se realiza el recorte de las imágenes utilizando las máscaras de segmentación para el cálculo del NDVI por especie.

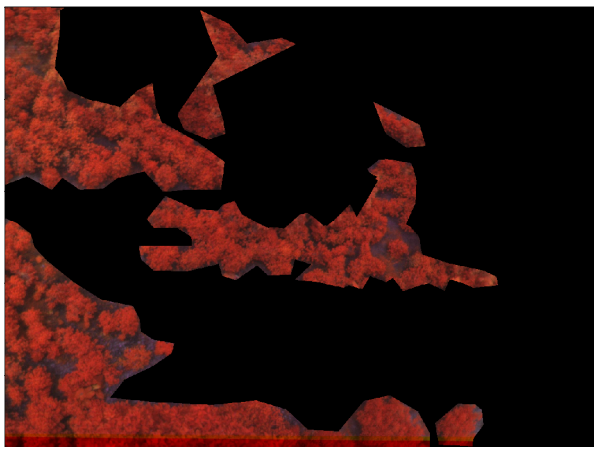


Figura 31. Segmentación de café.

Para calcular el valor de NDVI se toma la imagen multiespectral y se procesa de acuerdo a la ecuación 4, lo que resulta en una imagen en escala de grises (una capa de color). El valor medio de NDVI se calcula sumando la intensidad de cada uno de los píxeles mayores que cero y dividiendo el resultado entre el conteo de píxeles. Esta métrica puede

sesgar el cálculo a un valor inferior ya que cuenta píxeles como los del suelo u otras plantas. Utilizando la máscara de segmentación obtenida del modelo de aprendizaje profundo es posible ajustar el cálculo para que se realice para cada una de las especies vegetales. En la imagen 32 se tiene una sección cultivada con Musa Paradisiaca L (Plátano) y algunos arbustos. La corrección de valores de NDVI ayudan a aproximar mejor el estado de irrigación del cultivo.

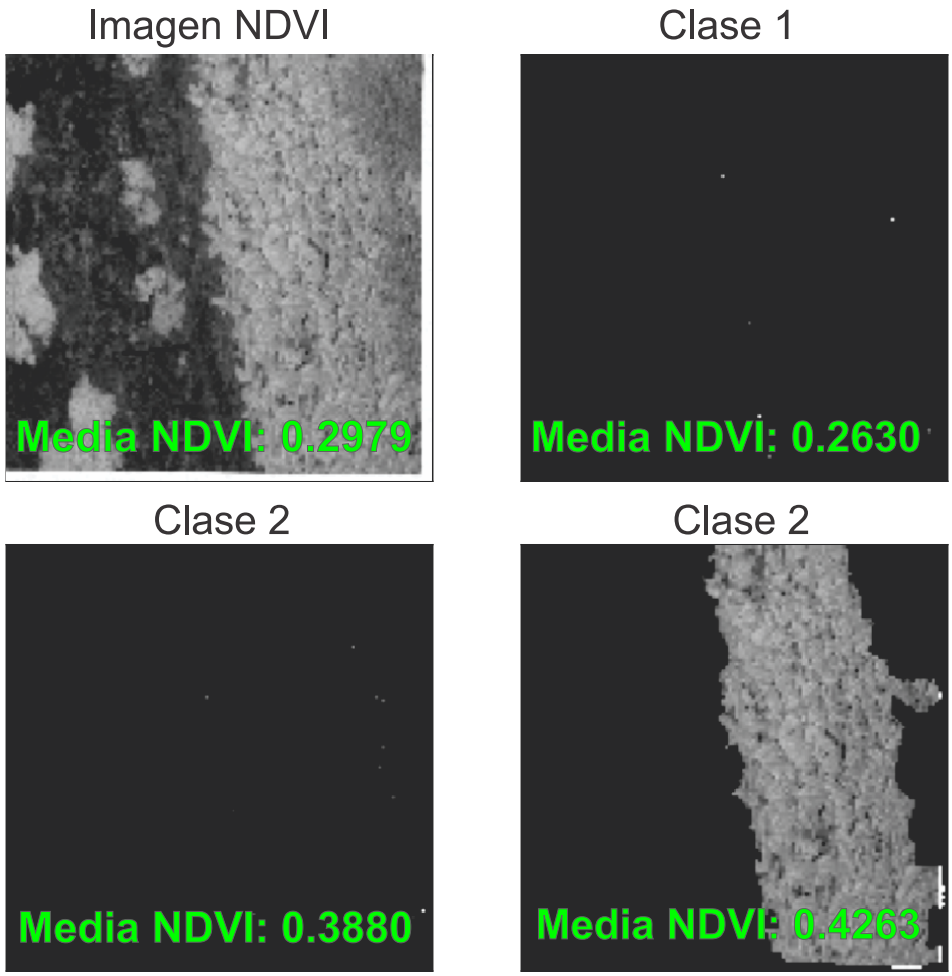


Figura 32. Ejemplo 1 de cálculo de NDVI ajustado por especie.

De forma similar se aprecia en la figura 33, la cual contiene los tres tipos de cultivos a detectar. En este caso, se aprecia una mejor aproximación del estado de salud para la Persea Americana y para Coffea Arabica. En el caso de la Musa Paradisciaca L,

los cultivos detectados son muy jóvenes por lo que predominan píxeles de suelo y la detección no es muy precisa, derivando en un valor de NDVI inferior al esperado.

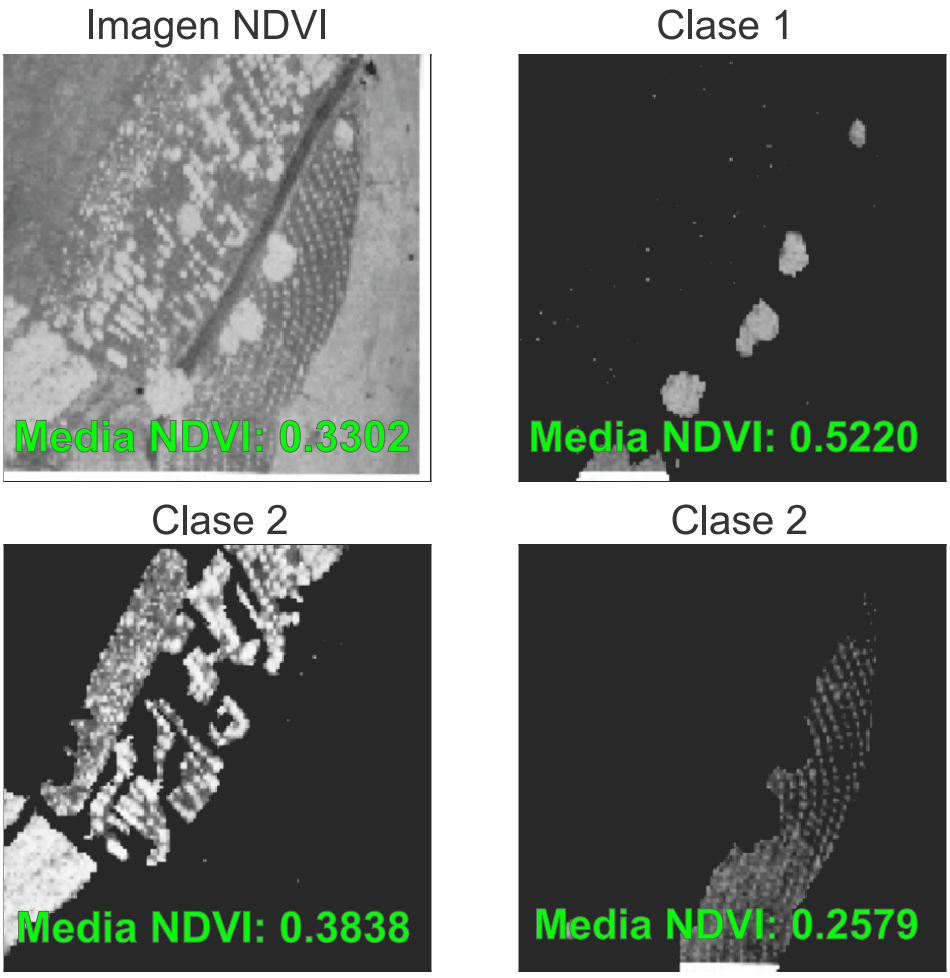


Figura 33. Ejemplo 2 de cálculo de NDVI ajustado por especie.

8. CONCLUSIONES Y RECOMENDACIONES

8.1. CONCLUSIONES

Con este trabajo se desarrolla un sistema de gestión de imágenes multi-espectrales de área amplia, el cual, sirve como una alternativa para el almacenamiento, la consulta, el etiquetado y la predicción de información que facilita el seguimiento de cultivos monitorizados con Drones y cámaras multi-espectrales.

El primer aporte de este trabajo es el desarrollo de una metodología de adquisición de imágenes multi-espectrales con un UAS. Por motivos de la pandemia declarada por la organización mundial para la salud hubo necesidad de cambiar el origen de los datos del proyecto, es decir, en lugar de emplear un UAS y dos cámaras diferentes se utilizó una base de datos adquirida por un UAS con una cámara de cinco lentes. Sin embargo, este cambio no afectó el desarrollo del primer objetivo específico ya que las imágenes tuvieron que ser tratadas y alineadas como se había previsto en el diseño metodológico. Los mismos procesos fueron aplicados al conjunto de imágenes para poder obtener la información sobre cultivos de Aguacate, Café y Plátano. Como resultado se generó una base de datos etiquetada que permite el entrenamiento de modelos de segmentación semántica para la detección de especies vegetales.

El segundo aporte desarrollado en este trabajo es el desarrollo de un sistema de gestión de imágenes multi-espectrales de área amplia. Para el desarrollo del sistema se recurrió a técnicas de organización de imágenes para consulta rápida acomodando las coordenadas en una curva de de llenado de espacio de Hilbert. Con la metodología descrita se redujo la complejidad de los algoritmos de búsqueda de imágenes por coordenadas espaciales. Adicionalmente, el sistema se puede consultar a través de un programa diseñado para la visualización y el manejo de imágenes multi-espectrales.

El tercer aporte de este trabajo es el entrenamiento de un modelo de aprendizaje profundo basado en la arquitectura *U-Net*. Este modelo fue entrenado con la base de datos generada y se desplegó sobre un sistema embebido. Específicamente, en la tarjeta de desarrollo Nvidia Jetson Nano™. Con el modelo entrenado es posible segmentar tres tipos diferentes de cultivos:*Musa Paradisiaca L* (Plátano), *Persea americana* (Aguacate) y *Coffea Arabica* (Café) en imágenes aéreas con el fin de ajustar el cálculo de índices de vegetación y dar un mejor diagnóstico del estado de irrigación de grandes extensiones de tierra cultivadas.

Finalmente, se realiza el cálculo de el índice de vegetación NDVI ajustado a las máscaras de segmentación semántica obtenidas con el modelo de Deep Learning. Esto permite acelerar el proceso de medición del estado de irrigación e incluso, a futuro permite realizar la detección en línea. Otra ventaja de calcular el NDVI por máscaras es la capacidad de discriminar el índice para cada especie. Debido a que la reflectancia entre especies varía, la estimación del estado de salud puede hacerse con una mejor precisión. Sin embargo, con la realización de este estudio, se ha encontrado que es necesario comparar las medidas de NDVI obtenidas con datos tomados en tierra utilizando un radiómetro (o fotómetro) para calibrar la medición de la cámara y validar que los ajustes de captura sean apropiados. Este tipo de calibración es necesaria para reducir el ruido y el sesgo que el tiempo de exposición de una cámara puede inducir en el cálculo de los índices de vegetación.

8.2. RECOMENDACIONES

Cuando se realiza la detección de vegetación a nivel de planta, el ruido generado durante el desplazamiento de los planos proyectivos de las imágenes, dificulta el análisis de el estado de salud de las plantas. Como recomendación se deben pre-procesar las

imágenes utilizando algoritmos de filtrado de paso de bajas y de conexión de componentes conexas. Con esto se puede mejorar el resultado de la estimación de los índices de vegetación a niveles cercanos al suelo e incluso en imágenes donde aparecen pocas plantas en una sola captura. Adicionalmente, para un cálculo ajustado a la reflectancia de las plantas en diferentes horas del día, se recomienda utilizar un radiómetro para medir en tierra la firma espectral de las plantas y así corregir el cálculo del NDVI.

Como trabajo futuro se propone la inclusión de variables medidas en tierra con el fin de mejorar la estimación del estado de irrigación. Si bien, se puede estimar un valor medio de NDVI con las máscaras detectadas y las imágenes multiespectrales aéreas, este valor debe ser calibrado por medio de la comparación de los datos obtenidos a través de imágenes y mediciones con una muestra estadística apropiada realizadas con un radiómetro que permita determinar la firma espectral de las hojas. Adicionalmente, se deben realizar diferentes tomas a lo largo del día ya que la radiación solar influye en la reflectancia y por lo tanto, en el estrés que pueden tener las plantas debido a la radiación. Se considera que incluyendo información de la hora y del tiempo es posible crear un modelo de aprendizaje profundo para refinar la estimación del NDVI incluyendo la información de la hora de captura y la radiación solar medida.

Debido a la topografía del Eje Cafetero, la cual predominantemente es región montañosa, se recomienda coordinar los vuelos desde locaciones altas, con el fin de mantener la línea de vista del UAS. Esto permite el control remoto sin riesgo de extravío o de colisión con montañas que pueden tener pendientes pronunciadas. Se recomienda estimar la altura máxima a la que debe realizarse la tarea de retorno a casa ya que en ciertos terrenos pueden existir cables de media o de alta tensión, postes o estructuras con las que el Drone pueda colisionar. Adicionalmente se recomienda el uso de múltiples baterías para la captura de datos a lo largo del día en diversas condiciones de radiación solar.

El desarrollo de sistemas que agrupan grandes cantidades de imágenes, así como la creación de bases de datos requiere de sistemas de computación distribuido en los cuales es necesario aplicar técnicas de transferencia de datos que permitan acelerar tareas de consulta y copiado. También se recomienda la implementación de certificados digitales y el uso de contraseñas seguras para evitar pérdidas de la información.

Respecto al entrenamiento de modelos de aprendizaje profundo, se observa que tienen buen potencial para tareas de agricultura de precisión. Experimentalmente se pudo observar que, con una base de datos pequeña se pudieron obtener resultados que mejoran los reportados por los autores en [34]. Sin embargo, las curvas de entrenamiento del modelo siguieren que se puede mejorar el rendimiento, pero que se requiere de una base de datos más grande para el entrenamiento.

Glosario

CRF *Conditional Random Field* El campo aleatorio condicional es un modelo estocástico utilizado para segmentación de datos y extracción de información de píxeles. Esta se desarrolla en [36].. 28

FTP File transfer protocol. 46

GPU Graphical Processor Unit.. 54

ISBI International Symposium on Biomedical Imaging: Esta organización lanza diferentes retos de segmentación de píxeles en imágenes con el fin de promover investigación en ingeniería aplicada en imágenes médicas. . 24

NDVI Índice normalizado diferencial de vegetación. Se calcula a partir de la luz capturada en una cámara en los canales rojo e infrarrojo como se aprecia en la ecuación (4). 13, 17, 20, 23, 32, 33, 65, 71, 80, 85

NIR Near infrarred channel. 39

PNG portable network graphics. 49, 67

RGB canales de color. 39, 46, 48

SIFT Scale Invariant Feature Transform: Es un algoritmo que permite extraer características en una imagen con la ventaja de que cada característica se preserva sin importar la escala o la rotación que presente una imagen.. 64

SOFTMAX Función que recibe como entrada un tensor y calcula la probabilidad de cada uno de sus elementos de pertenecer a alguna clase k predefinida. 26

SQL structured query language. 46, 48

UAS *Unmanned Aerial System:* Todo sistema que incluye vehículos aéreos no tripulados y sus elementos de mando y control remoto.. 11, 14, 53–55, 63, 85

VRAM Es la memória gráfica de acceso aleatorio y se utiliza para almacenar los datos de núcleos de procesadores gráficos o GPU.. 53, 54

BIBLIOGRAFÍA

[1] M. Nery, R. Santos, W. Santos, V. Lourenco, and M. Moreno. Facing digital agriculture challenges with knowledge engineering. In *2018 First International Conference on Artificial Intelligence for Industries (AI4I)*, pages 118–119, Sep. 2018. 1, 1.1

[2] Yanbo Huang, Zhong xin CHEN, Tao YU, Xiang zhi HUANG, and Xing fa GU. Agricultural remote sensing big data: Management and applications. *Journal of Integrative Agriculture*, 17(9):1915 – 1931, 2018. 1, 1.1, 1.2

[3] Philip K Thornton, Patricia Kristjanson, Wiebke Förch, Carlos Barahona, Laura Cramer, and Sonali Pradhan. Is agricultural adaptation to global change in lower-income countries on track to meet the future food production challenge? *Global Environmental Change*, 52:37 – 48, 2018. 1, 6

[4] Gong Cheng and Junwei Han. A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117:11–28, 2016. 1

[5] Gonzalo Pajares. Overview and current status of remote sensing applications based on unmanned aerial vehicles (uavs). *Photogrammetric Engineering and Remote Sensing*, 81(4):281 – 329, 2015. 1, 1.1, 1.2, 2.6.2

[6] Fengsong Pei, Changjiang Wu, Xiaoping Liu, Xia Li, Kuiqi Yang, Yi Zhou, Kun Wang, Li Xu, and Gengrui Xia. Monitoring the vegetation activity in china using vegetation health indices. *Agricultural and Forest Meteorology*, 248:215 – 227, 2018. 1, 1.1

[7] W. L. da Silva, R. R. V. Gonçalves, A. S. Siqueira, J. Zullo, and F. A. M. G. Neto. Feature extraction for ndvi avhrr/noaa time series classification. In *2011 6th International Workshop on the Analysis of Multi-temporal Remote Sensing Images (Multi-Temp)*, pages 233–236, July 2011. 1, 1.1, 2.1, 2.6.1

[8] J. Huang, H. Wang, Q. Dai, and D. Han. Analysis of ndvi data for crop identification and yield estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(11):4374–4384, Nov 2014. 1

[9] M. R. Mobasheri A, M. Chahardoli B, J. Jokar C, and M. Farajzadeh D. Sugarcane phenological date estimation using broad-band digital cameras. 1

[10] Andreas Kamilaris and Francesc X. Prenafeta-Boldú. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147:70 – 90, 2018. 1, 1.1, 1.2, 5

[11] J. Natividade, J. Prado, and L. Marques. Low-cost multi-spectral vegetation classification using an unmanned aerial vehicle. In *2017 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pages 336–342, April 2017. 1.1

[12] Larry K. B. Li David D. W. Ren, Siddhant Tripathi. Low-cost multispectral imaging for remote sensing of lettuce health. *Journal of Applied Remote Sensing*, 11:11 – 11 – 13, 2017. 1.1

[13] Hoffmann M. Zarezadeh AA. Bobda C. Dworak V. Selbeck, J. Dammer K-H. Strategy for the development of a smart ndvi camera system for outdoor plant detection and agricultural embedded systems. *Sensors (Basel, Switerland)*, 2013. 1.1, 2.6.2

[14] Zengyuan Li, Xiaohong Li, Erxue Chen, and Shiming Li. A method integrating gf-1 multi-spectral and modis multi-temporal ndvi data for forest land cover clas-

sification. In *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, pages 3742–3745. IEEE, 2016. 1.1

[15] E. Raymond Hunt, Michel Cavigelli, Craig S. T. Daughtry, James E. Mcmurtrey, and Charles L. Walthall. Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. *Precision Agriculture*, 6(4):359–378, Aug 2005. 1.1

[16] Stefanos Georganos, Abdulhakim M. Abdi, David E. Tenenbaum, and Stamatis Kalogirou. Examining the ndvi-rainfall relationship in the semi-arid sahel using geographically weighted regression. *Journal of Arid Environments*, 146:64 – 74, 2017. 1.1

[17] Valentine Lebourgeois, Agnès Bégué, Sylvain Labbé, Benjamin Mallavan, Laurent Prévot, and Bruno Roux. Can commercial digital cameras be used as multispectral sensors? a crop monitoring test. *Sensors*, 8(11):7300–7322, 2008. 1.1, 2.1

[18] R. B. Myneni, F. G. Hall, P. J. Sellers, and A. L. Marshak. The interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing*, 33(2):481–486, March 1995. 1.1, 1.2, 2.6.2

[19] Rangel Daroya and Manuel Ramos. Ndvi image extraction of an agricultural land using an autonomous quadcopter with a filter-modified camera. In *Control System, Computing and Engineering (ICCSCE), 2017 7th IEEE International Conference on*, pages 110–114. IEEE, 2017. 1.1, 1.2, 2.1, 2.6.1, 2.6.2, 3.2, 6, 6

[20] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, Nov 2000. 1.1, 2.1, 3.2, 7.1

[21] G. Rabatel, N. Gorretta, and S. Labbé. Getting NDVI spectral bands from a single standard RGB digital camera: a methodological approach. In *14th Conference of the Spanish Association for Artificial Intelligence, CAEPIA 2011*, page 10 p., La Laguna, Spain, November 2011. Springer-Verlag. 1.1, 1.2, 2.1

[22] ministerio de agricultura y desarrollo rural de Colombia Estadísticas. Reporte tercer censo nacional agropecuario, 2016. 1.2

[23] Diego Inácio Patrício and Rafael Rieder. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and Electronics in Agriculture*, 153:69 – 81, 2018. 1.2

[24] B. José, M. Nicolás, C. Danilo, and A. Eduardo. Multispectral ndvi aerial image system for vegetation analysis by using a consumer camera. In *2014 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*, pages 1–6, Nov 2014. 1.2, 2.6.1, 2.6.1, 6

[25] O. A. Martinez S., A. Franco, and G. A. Holguin. Aerial wide-area image storage and retrieval for deep learning training. In *Avances y experiencias innovadoras en computación e informática*, volume 1, pages 25–40, 2020. 2.1

[26] Andreas Pawelke and Anoush Rima Tatevossian. Data philanthropy: Where are we now. *United Nations Global Pulse Blog*, 2013. 2.2

[27] Venkat N. Gudivada, Srini Ramaswamy, and Seshadri Srinivasan. 7 - data management issues in cyber-physical systems. In Lipika Deka and Mashrur Chowdhury, editors, *Transportation Cyber-Physical Systems*, pages 173 – 200. Elsevier, 2018. 2.2

[28] T. Sharma, V. Shokeen, and S. Mathur. Distributed processing of satellite images on hadoop to generate normalized difference vegetation index images. In *2017*

International Conference on Computing, Communication, Control and Automation (ICCUBEA), pages 1–5, Aug 2017. 2.2.2

[29] Konstantin Shvachko, Hairong Kuang, Sanjay Radia, Robert Chansler, et al. The hadoop distributed file system. In *MSST*, volume 10, pages 1–10, 2010. 2.2.2

[30] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. 2, 2.5, 2.5, 2.5, 5

[31] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. 2.3

[32] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015. 2.3

[33] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 2.3

[34] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 2.4, 1, 2.4, 2.5, 5, 7.2.1, 8.2

[35] Muhammad Hamza Asad and Abdul Bais. Weed density estimation using semantic segmentation. In Joel Janek Dabrowski, Ashfaqur Rahman, and Manoranjan Paul, editors, *Image and Video Technology*, pages 162–171, Cham, 2020. Springer International Publishing. 2.4, 5

[36] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017. 2.5, 5, 6, 5, 8.2

[37] J. Praveen Kumar and S. Domnic. Image based leaf segmentation and counting in rosette plants. *Information Processing in Agriculture*, 6(2):233 – 246, 2019. 2.5

[38] R. Barth, J. IJsselmuiden, J. Hemming, and E.J. Van Henten. Data synthesis methods for semantic segmentation in agriculture: A capsicum annuum dataset. *Computers and Electronics in Agriculture*, 144:284 – 296, 2018. 2.5, 5

[39] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014. 2.5, 5, 7.2.1

[40] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9):1904–1916, 2015. 2.5, 4

[41] Petra Bosilj, Tom Duckett, and Grzegorz Cielniak. Connected attribute morphology for unified vegetation segmentation and classification in precision agriculture. *Computers in Industry*, 98:226 – 240, 2018. 2.5

[42] Nived Chebrolu, Philipp Lottes, Alexander Schaefer, Wera Winterhalter, Wolfram Burgard, and Cyrill Stachniss. Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields. *The International Journal of Robotics Research*, 36(10):1045–1052, 2017. 2.5

[43] A. Milioto, P. Lottes, and C. Stachniss. Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in cnns.

In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2229–2235, May 2018. 2.5

[44] A. Peña, I. Bonet, D. Manzur, M. Góngora, and F. Caraffini. Validation of convolutional layers in deep learning models to identify patterns in multispectral images. In *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6, June 2019. 2.6.1

[45] L. Geng, M. Ma, W. Yu, X. Wang, and S. Jia. Validation of the modis ndvi products in different land-use types using in situ measurements in the heihe river basin. *IEEE Geoscience and Remote Sensing Letters*, 11(9):1649–1653, 2014. 2.6.1, 6

[46] M. Abuzar, D. Whitfield, A. McAllister, G. Lamb, K. Sheffield, and M. O'Connell. Satellite remote sensing of crop water use in an irrigation area of south-east australia. In *2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS*, pages 3269–3272, 2013. 2.6.1

[47] Anjin Chang, Jinha Jung, Murilo M. Maeda, and Juan Landivar. Crop height monitoring with digital imagery from unmanned aerial system (uas). *Computers and Electronics in Agriculture*, 141:232 – 237, 2017. 2.6.2

[48] Muhammad Adeel Hassan, Mengjiao Yang, Awais Rasheed, Guijun Yang, Matthew Reynolds, Xianchun Xia, Yonggui Xiao, and Zhonghu He. A rapid monitoring of ndvi across the wheat growth cycle for grain yield prediction using a multi-spectral uav platform. *Plant Science*, 282:95 – 103, 2019. The 4th International Plant Phenotyping Symposium. 2.6.2

[49] A. Ramanath, S. Muthusrinivasan, Y. Xie, S. Shekhar, and B. Ramachandra. Ndvi versus cnn features in deep learning for land cover clasification of aerial images. In

IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pages 6483–6486, July 2019. 2.6.2

[50] Hao Gan, Won Suk Lee, and Victor Alchanatis. A photogrammetry-based image registration method for multi-camera systems – with applications in images of a tree crop. *Biosystems Engineering*, 174:89 – 106, 2018. 3.2

[51] Guo-Rong Cai, Pierre-Marc Jodoin, Shao-Zi Li, Yun-Dong Wu, Song-Zhi Su, and Zhen-Kun Huang. Perspective-sift: An efficient tool for low-altitude remote sensing image registration. *Signal Processing*, 93(11):3088 – 3110, 2013. 3.2

[52] Kentaro Wada. labelme: Image Polygonal Annotation with Python. https://github.com/wkentaro/labelme, 2016. 4, 4.4, 10, 7.1

[53] T. Akgun, Y. Altunbasak, and R. M. Mersereau. Super-resolution reconstruction of hyperspectral images. *IEEE Transactions on Image Processing*, 14(11):1860–1875, Nov 2005. 4.2, 4.2

[54] Mohamed F. Mokbel and Walid G. Aref. *Space-Filling Curves*, pages 1068–1072. Springer US, Boston, MA, 2008. 8

[55] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 4.4

[56] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010. 4.4

[57] M. Fawakherji, A. Youssef, D. Bloisi, A. Pretto, and D. Nardi. Crop and weeds classification for precision agriculture using context-independent pixel-wise segmentation. In *2019 Third IEEE International Conference on Robotic Computing (IRC)*, pages 146–152, 2019. 5

[58] C. Liu, H. Li, A. Su, S. Chen, and W. Li. Identification and grading of maize drought on rgb images of uav based on improved u-net. *IEEE Geoscience and Remote Sensing Letters*, pages 1–5, 2020. 5

[59] Ute Schuppler, PH He, Peter John, and Rana Munns. Effect of water stress on cell division and cell-division-cycle 2-like cell-cycle kinase activity in wheat leaves. *Plant physiology*, 117:667–78, 07 1998. 6

[60] A. Wahid and T. J. Close. Expression of dehydrins under heat stress and their relationship with water relations of sugarcane leaves. *Biologia plantarum*, 51(1):104–109, 2007. 6

[61] M. Abuzar, D. Whitfield, A. McAllister, G. Lamb, K. Sheffield, and M. O'Connell. Satellite remote sensing of crop water use in an irrigation area of south-east australia. In *2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS*, pages 3269–3272, 2013. 6

[62] M. P. Arakeri, B. P. Vijaya Kumar, S. Barsaiya, and H. V. Sairam. Computer vision based robotic weed control system for precision agriculture. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1201–1205, 2017. 6

[63] E. Sanchez Franco. Metodología para la captura automática y sincronizada de imágenes aéreas multiespectrales. In *Universidad Tecnológica de Pereira.*, 2019. 7.1

[64] Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the variance of the adaptive learning rate and beyond, 2019. 7.2.1

[65] Luke Taylor and Geoff Nitschke. Improving deep learning using generic data augmentation. *arXiv preprint arXiv:1708.06020*, 2017. 7.2.1